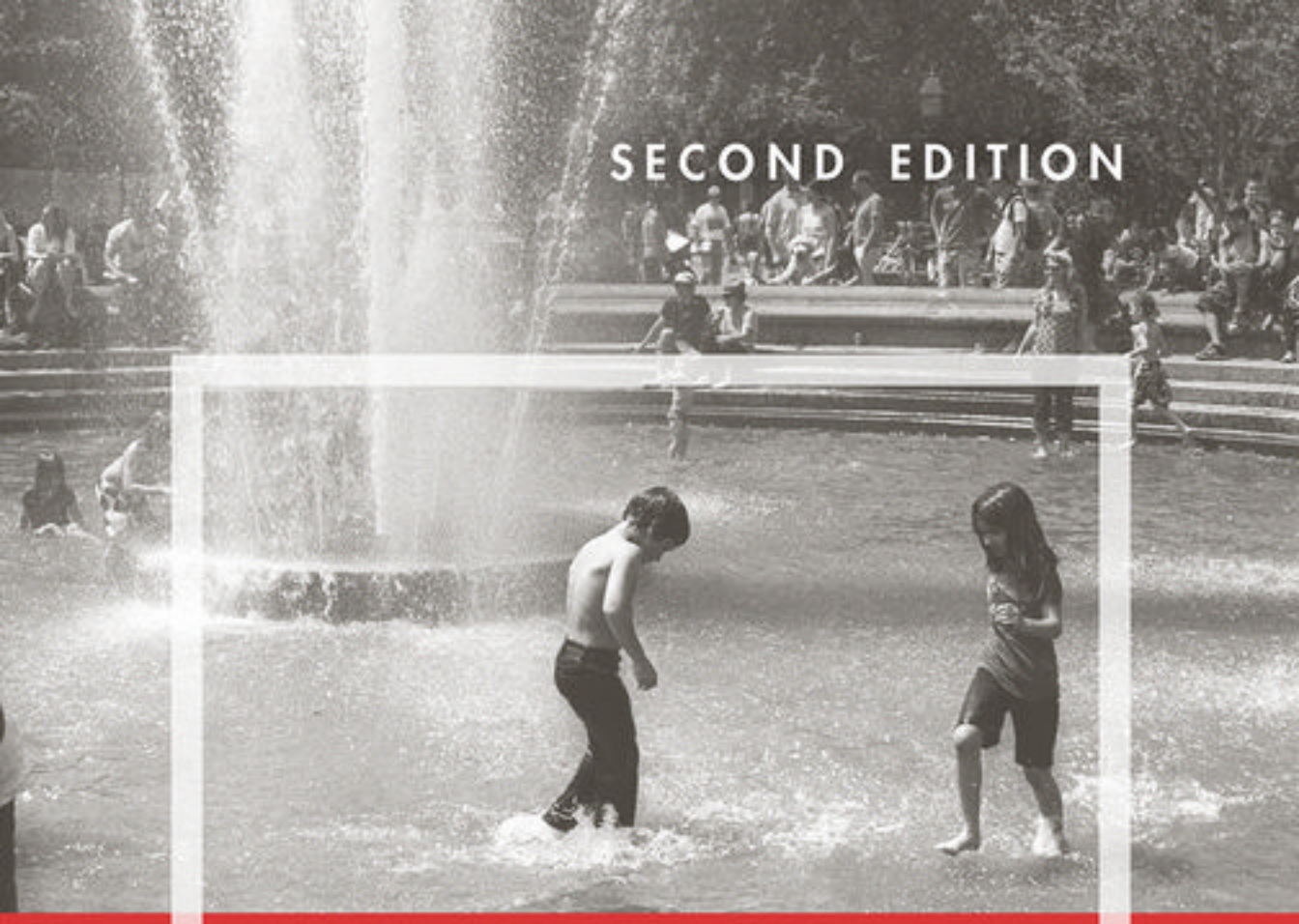


SECOND EDITION

SOCIAL EPIDEMIOLOGY

EDITED BY

Lisa F. Berkman

Ichiro Kawachi

M. Maria Glymour

OXFORD

SOCIAL EPIDEMIOLOGY

SOCIAL EPIDEMIOLOGY

SECOND EDITION

Edited by

Lisa F. Berkman

Ichiro Kawachi

M. Maria Glymour

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press is a department of the University of
Oxford. It furthers the University's objective of excellence in research,
scholarship, and education by publishing worldwide.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trademark of Oxford University Press
in the UK and certain other countries.

Published in the United States of America by
Oxford University Press
198 Madison Avenue, New York, NY 10016

Library of Congress Cataloging-in-Publication Data
Social epidemiology / edited by Lisa F. Berkman, Ichiro Kawachi, M. Maria Glymour. — Second edition.
p. ; cm.
ISBN 978–0–19–537790–3 (hbk.) — ISBN 978–0–19–939533–0 (pbk.) (alk. paper)
I. Berkman, Lisa F., editor. II. Kawachi, Ichiro, editor. III. Glymour, M. Maria, editor.
[DNLM: 1. Social Medicine. 2. Epidemiology. 3. Health Status Disparities. 4. Public Policy. WA 31]
RA418
306.4′61—dc23 2014006489

9 8 7 6 5 4 3 2 1
Printed in the United States of America
on acid-free paper

CONTENTS

FOREWORD TO SECOND EDITION OF *SOCIAL EPIDEMIOLOGY*

I was very pleased when the first edition of Social Epidemiology appeared in the year 2000. That book established a new subdiscipline in the field of Epidemiology. It was the first textbook ever written that pulled together, in one place, the evidence that was beginning to accumulate showing a connection between social factors and health. This was an especially interesting phenomenon for me: when I began to work on this topic in 1958, there was almost nothing known about this issue. There were very few of us studying social factors at that time. I remember such people as Saxon Graham, John Cassel, Sol Levine, and Leo Reeder but none of us could have imagined a real textbook ever being written about social factors.

Here we are 14 years later and a *second* edition of Social Epidemiology has now been published! During these 14 years, several other major textbooks dealing with Social Epidemiology have also been published and courses in this field have been established in almost every School of Public Health in the world. The reason for this explosion of interest is not difficult to explain. First, it is becoming increasingly clear that the individual risk factors we all know about (serum cholesterol, blood pressure, cigarette smoking, obesity, physical inactivity, poor diet) account for only a relatively small fraction of the diseases that occur. Second, our efforts to help people lower those individual risks are only modestly successful. But third, and most important, even if everyone succeeded in lowering their risk profiles, new people would continue to enter the at-risk population forever because we rarely target those social forces in society that cause the problem in the first place.

The maturation of social epidemiology is of great importance because it provides several perspectives on epidemiological research that are crucial to its mission. Two of these perspectives are of special significance. One of them involves a much-needed focus on family, neighborhood, community, and the social group. The second perspective involves a more appropriate way to study risk factors and diseases that can fundamentally change our approach to the concepts of etiology and intervention. These are nontrivial contributions and each deserves to be considered in more detail.

Let us consider the first perspective: the focus on the group. A major purpose of epidemiology is to contribute information relevant to the prevention of disease and the promotion of health. To accomplish this goal, epidemiologists study the distribution of disease in populations and attempt to identify the factors that explain that distribution. As is evidenced throughout this book, the special perspective of social epidemiology brings to the field of epidemiology more than just an additional set of factors that can be studied. This perspective emphasizes the fact that health and disease are influenced by factors not only at the individual level but also at the group or community level. This approach is in startling contrast to many epidemiologic studies that focus on individuals and individual risk factors. Thus, many so-called "community" studies in epidemiology really seem to consist of careful descriptions of individual behaviors and characteristics as these are related to the occurrence of disease. This research might more properly be seen as clinical research in large groups of people.

I have always considered the work of Émile Durkheim on suicide as providing a remarkable and valuable illustration of the importance of a social epidemiologic approach. Durkheim demonstrated the importance of the social environment by studying one of the most individual and intimate behaviors imaginable—suicide. In his work, Durkheim noted that suicide rates in countries and groups exhibit a patterned regularity over time, even though individuals in these groups come and go. If suicide is a product of anguishing intimate and deeply personal problems, it is puzzling to see that rates of suicide in these groups remain higher or lower even though individuals move in and out of the groups. The answer, Durkheim suggested, was to be found in the social environment of these groups. These social factors in the environment would not, of course, determine which individuals in the group would commit suicide but they would help to explain group differences in the rate over time.

The perspective that Durkheim offered was to see that the health and well-being of a *community* were affected by the social milieu within which people lived. When one looks only for risk factors at the individual level, our interventions inevitably focus on individual behavior. The problem with this approach is that even if these interventions were completely successful, new people would continue to enter the at-risk population at an unaffected rate since, as I noted earlier, we have done nothing to influence those forces in the community that caused the problem in the first place. So in this book we have a new opportunity to think more clearly and creatively about the real meaning of the social environment. This focus hopefully will encourage us to approach the issues of etiology *and* intervention in a fresh and more meaningful way.

The second perspective that social epidemiology helps us address is how we classify disease. Epidemiology is failing to solve the main problem it is intended to address. A major task of epidemiology is to identify risk factors for disease. The failure of the field to successfully accomplish this objective for many chronic diseases can be illustrated by the work that has been done on coronary heart disease—the number one cause of death in the industrialized countries of the world. Coronary heart disease has been studied by epidemiologists since the early 1950s in the most aggressive, well-financed manner the world has ever seen. During these years of massive worldwide effort, a large number of important risk factors have been identified. The three that everyone agrees on are cigarette smoking, high blood pressure, and high serum cholesterol. Dozens of other risk factors have been proposed, but not everyone agrees about them: obesity, physical inactivity, diabetes, blood lipid and clotting factors, stress, and various hormone factors. Nevertheless, when *all* of these risk factors are considered together, they explain only about 40% of the coronary heart disease that occurs.

How is it possible that after 50 years of effort, all of the risk factors we know about, combined, account for less than half of the disease that is identified? Is it possible that we have somehow missed one or two crucial factors? This is, of course, conceivable, but the relative risk of these missing risk factors would have to be enormous to account for the other 60% of the coronary heart disease that occurs. It seems not very likely that we would have missed one or two risk factors of such enormous power and importance. Is this a problem limited to heart disease? It turns out that the problem is even more challenging for other diseases. So, clearly, we need to do some re-thinking about our approach to disease etiology.

An early pioneer in the field of social epidemiology, John Cassel, suggested an explanation for this problem in a classic paper he wrote in 1976, just before his death. In this article, Cassel noted that a wide variety of disease outcomes were associated with similar circumstances. For example, he cited the remarkably similar set of risk factors that characterized people who developed tuberculosis or schizophrenia, people who became alcoholics, and those who were victims of multiple accidents or who committed suicide. Cassel also noted that this phenomenon generally had escaped comment, because, he suggested, investigators usually are "concerned with only one clinical entity, so that features common to multiple disease manifestations have tended to be overlooked."

We in epidemiology have adopted a disease classification scheme that is based on a clinical approach to disease. This approach, of course, is yet another legacy of our focus on the individual instead of the group. There is no question that this clinical approach is of value in diagnosing and treating disease in sick people, but it is not as useful if our goal is to *prevent* disease. Infectious disease epidemiologists of an earlier era solved this problem by classifying disease in a far more appropriate and useful way. They studied water-borne diseases, air-borne diseases, food-borne diseases, and vector-borne diseases. This classification scheme helped us think about interventions in a more effective way by targeting those elements of the environment that were responsible for the disease problem. We have not yet developed a comparable set of categories for the study of such noninfectious diseases and conditions as heart disease, cancer, injuries, and suicide.

Part of the difficulty here is that our major source of research funds, the National Institutes of Health (NIH), is so fundamentally organized around the clinical model of disease. If you sent a research grant proposal to NIH to study poverty diseases, they would not know to which disease-specific institute it should be sent for review. The same problem would exist for a proposal on smoking diseases. And nutritional deficiency diseases. And stress diseases. That we are forced to think about these issues in disease-specific terms weakens our ability to think about social forces that transcend particular organ damage.

This book on social epidemiology offers a fresh approach to the problem. Not one chapter is organized around a clinical disease. Instead, the book focuses on those major social forces and concepts that influence the occurrence of disease and that perhaps can be used to think more creatively about new ways to classify disease and new ways to think about interventions. This is a major innovation and contribution to our thinking.

There is another fresh feature to this book. While the 2000 edition covered emerging research on inequalities, neighborhood, community, work, and family, this new volume includes discussions of yet another set of important topics that are only now beginning to be recognized. The book reviews the impact on health of various public policies, including labor, educational, family, and other policies. These issues have a profound impact on the health of our nation, and it is good

that we can now begin to recognize that fact. This new edition also provides a lifecourse perspective on everything that is discussed. It turns out that virtually every risk factor we think about affects us from our earliest years to our last days and we need to take that into account in our thinking. Major consideration is now also given in this edition to thinking about how risk factors get into the body to affect health. We are no longer content to simply observe associations between social factors and health; we want to see how these social factors influence our biology. A better understanding of how the social environment is embedded in our biological functioning opens whole new avenues for study and intervention.

While thinking about what I wanted to write for this foreword, I glanced at a volume that Leo Reeder and I edited in 1967 called *Social Factors and Cardiovascular Disease*. Contributors to that book were some of the most distinguished scholars in the field at the time and we invited all of them to a conference to discuss our new field. The entire group consisted of only 28 people! When I compared the state of our science 47 years ago to what is contained in the present volume, I was in awe. While the material in the 1967 book is primitive compared to our present knowledge, it is clear that the work we were doing in those early years was headed in the right direction. I hope that 47 years from now, in the year 2061, someone will be able to write a foreword to a new volume on social epidemiology, and I hope that a similar leap will be seen in creativity, methodology, and sophistication. But I hope it will be clear also that this current volume laid an outstanding foundation for that future achievement. Our ability to progress in health promotion and disease prevention depends on it.

S. Leonard Syme, PhD
Professor Emeritus, Epidemiology and Community Health
School of Public Health
University of California, Berkeley

PREFACE

> Although we set out primarily to study reality, it does not follow that we do not wish to improve it; we should judge our researches to have no worth at all if they were to have only speculative interest. If we separate carefully the theoretical from the practical problems, it is not to the neglect of the latter, but on the contrary, to be in a better position to solve them.
>
> Emile Durkheim, *The Division of Labor*

This is a bold new book. While the title remains the same, almost all of the material in the book is new. This is an ode to the success of social epidemiology as an approach to understanding the fundamental determinants of health. When we published the first edition of this book in the late 1990s there were handfuls of papers scattered across journals to substantiate the role that our social world plays in shaping population patterns of health and illness. There is now so much new work that each of the chapters could be a book in itself. Where once there were 6 or 8 studies on a topic there are now meta-analyses of hundreds of papers from around the world.

Over the last 40 years, there has been an explosion of interest in how society and different forms of social organization influence health and well-being, along with a recognition that such understanding can guide powerful strategies to reducing inequalities and improving population health. The field of social epidemiology has emerged during this time, drawing heavily on public health work done during the early part of this century by Frost, Goldberger, and Sydenstricker; on work on stress by Cannon and Selye; and on the blossoming fields of medical sociology and health psychology. Where epidemiology was once comfortable in assessing only the role of the physical environment in determining health outcomes, we now have the tools with which to assess the impact of the social environment. This volume represents a second major attempt to describe the new methods and theories that have developed over the last 15 years and to review the vast

empirical evidence in a number of social domains. Our aim is to provide the reader, from graduate student to active investigator, with a guide to the major social conditions of importance and to new approaches in public policy, demography, economics, social sciences, and of course in epidemiology. Contributors have generally provided both theoretical and methodological overviews of their respective areas that should help investigators launch their own research, building on the most up-to-date information available.

This second edition is substantially new in its orientation to the impact of public policies and natural and randomized experiments. There are several completely new chapters in the book that cover labor policies and economic conditions, evaluation of the health impacts of broad policies drawing on natural experiments in the policy arena, health economics, and implementation of social epidemiology in agenda-setting in policy. In addition, virtually all chapters draw on evaluation of policies that have altered the landscape with regard to discrimination, education, and working conditions. It is a new world for epidemiology: one in which scientists not only observe the world using classical approaches, but also design and evaluate interventions to improve health. In many ways, this is an alliance that was shared by early public health workers in the nineteenth and twentieth centuries when both practice and research were integrated in public health.

The book is organized into 15 chapters. After a foreword by S. L. Syme that embeds *Social Epidemiology* within a historical context, our introduction on the history of social epidemiology deals with overarching issues in the field from its beginnings to the current day. The first group of chapters deals with socioeconomic status, discrimination, and income inequalities in health. It begins with a chapter by Glymour, Avendano, and Kawachi that covers evidence on the link between individual socioeconomic status and health. The chapter now covers natural experiments that evaluate the role of education in long-run health outcomes. Krieger then explores the role of discrimination largely related to race and ethnicity but also to gender, sexual orientation, and age. Finally, Kawachi and Subramanian review the growing literature on area-based socioeconomic inequalities in health. Together, these three chapters provide the latest theories and evidence on the pervasive impact that socioeconomic position and discrimination have on health outcomes. They contribute to our understanding of the racial and ethnic disparities in health that are so prevalent in the United States by analyzing the social conditions that underlie them.

The next two chapters examine the work environment and the labor market in relation to health status. Berkman, Kawachi, and Theorell review the development of major concepts in organizational conditions related to work including job strain, effort/reward imbalance, work-family conflict, shift work, and schedule control. Avendano and Berkman discuss the influence of macro-economic conditions related to job insecurity, unemployment, and recessions. They then consider labor policies that impact health including those related to retirement, maternal leave, and unemployment. These two chapters bring us up to date on current theory, measures, and methodologic problems in the study of work and health.

The role of community and social relations in health is the theme of the third set of chapters. Berkman and Krishna tie together theoretical approaches and evidence concerning the effect on health of social integration, social networks, and social support. This chapter incorporates the major efforts in social network analysis and the ways they have been integrated into social epidemiology. Then Kawachi and Berkman review the evidence on social capital as it relates to health. As in the first section, here, too, area-based and individual-level assessments are discussed.

The fourth section reviews the dynamic state of research on how affective states are associated with health outcomes, especially cardiovascular disease. Kubzansky, Winning, and Kawachi review the data related to emotional states, including positive and negative dimensions of emotional health. These psychological states are important in their own right and as pathways that mediate the influence of social circumstances on health. Lifecourse issues are covered here as they relate to the development of emotional health.

The final set of chapters covers a number of issues that are central to social epidemiology and that require a truly multidisciplinary perspective. Okechukwu, Davison, and Emmons discuss the social context of health-promoting and health-damaging behaviors and how behavioral interventions might benefit from a deeper integration of social organization into behavioral interventions. Glass and Berkman present new psychosocial models of intervention where the aim is to modify the social milieu as well as the psychological condition of individuals and groups. Workplace interventions that aim to improve health behaviors as well as those that alter organizational conditions are discussed in detail in both of these chapters. We have added two completely new chapters on policies as tools for research and translation in epidemiology by Glymour and on behavioral economics by Kawachi. Both of these chapters reflect the extent to which the fields of economics and public policy have fundamentally altered how we evaluate health impacts of policies and how we can change behaviors to improve health. These areas are critical to social epidemiology since we have little hope of improving health if we cannot influence public and private sector policy and behavioral decisions. We move on to a detailed review of hypothesized biological pathways linking social conditions to health. The chapter, written by Kubzansky, Seeman, and Glymour, represents some of the most innovative thinking—and speculation—about the biological embedding of social experiences. In this chapter, the authors review relatively well-investigated pathophysiologic mechanisms, and outline pathways about which much less is known in this rapidly unfolding area. Finally, *Social Epidemiology* concludes with a chapter by Marmot and Allen on the implementation of social determinants of health into a global policy perspective. The underlying theme of the volume is that in order to improve health, we must move beyond traditional medical or health care policy to understand the impact on the health of the public by social organization, social structure, and the policies that shape them.

L.F.B. I.K., M.M.G.
Cambridge, Mass.
March 2014

ACKNOWLEDGMENTS

Our new volume of *Social Epidemiology* brings on a fresh set of thanks and acknowledgments for those who have enabled this field to flourish over the last decade. First, both Ichiro and I would like to acknowledge the enormous contribution that Maria Glymour has made to this edition. Enlisting her as a co-editor hardly does justice to the impact that she has had on this field over the last decade. While it is true that we stand on the shoulders of those who have come before us, it is the energy and insights of the next generation of scholars that create a dynamic field. In this volume, we would like to start by extending our thanks to the new authors who have changed the field of social epidemiology. This generation includes Maria Glymour, Mauricio Avendano-Pabon, Subu Subramanian, Cassandra Okechukwu, and Kirsten Davison. We see on the horizon yet another generation of young scholars, including Aditi Krishna, Amy Ehntholt, Ashley Winning, Amii Kress, and Jessica Allen, who each made truly outstanding contributions to this volume; we are confident that we will see them as emerging social epidemiologists in the next decade. Of course, the field would not be what it is today without the contributions of our senior authors, Nancy Krieger, Laura Kubzansky, Teresa Seeman, Thomas Glass, Tores Theorell, and Michael Marmot, who have done so much to shape this young field, and who have simultaneously trained, engaged, and inspired successive generations of scholars. In any network analysis or genealogy, S. Leonard Syme would stand at the network hub or origins of this book. Having trained Berkman, Marmot, Krieger, and Seeman, and worked closely with Theorell and Kawachi, Len has single-handedly served to shape this field. As the network branches out—especially to Harvard—Berkman, Kawachi, and Krieger have taught many of the next generation of scholars, and consequently, new nodes in a number of US, UK, and European universities now have flourishing programs. This is our "ego-centric" network view of social epidemiology.

Now, for the socio-centric view. A socio-centric view of social epidemiology reveals numerous network hubs throughout the United States and around the world. The field would not be the same without the Robert Wood Johnson Foundation, and their launching of the Scholars in Health Society Fellowship. We thank Steve Schroeder and Risa Lavizzo-Mourey for their vision of what the field of society and health would look like if it incorporated a truly multidisciplinary range of scientists from sociology and psychology to epidemiology, history, and the biological sciences.

With Centers at University of Michigan (Kaplan, Diez Roux, House, Morenoff), Columbia (Link, Bearman), UC Berkeley/UCSF (Adler, Catalano, Syme), University of Wisconsin (Kindig, Robert, Mullahy), University of Pennsylvania (Aronowitz, Asch) and Harvard, the Foundation has for over a decade enabled scholars to invest in innovative thinking about the social determinants of population health. At Harvard, our community has thrived because of the interdisciplinary adventures spurred by Allan Brandt, Charles Rosenberg, Sandy Jencks, and all of our social epidemiology colleagues. This network of senior leaders and the new generation of scholars will continue to initiate novel approaches to this field for decades to come. We thank Chris Bachrach and Jo Ivey Bufford for their leadership of the program and Harvey Fineberg for his vision and guidance as well as his leadership over the years at Harvard, both as Dean of HSPH and as Provost. Their contributions have been fundamental to the growth of the fields of population health and social epidemiology. Internationally, a network of social epidemiologists at Erasmus in the Netherlands (Mackenbach), in Sweden at the Karolinska (Theorell, Orth-Gomer), CHESS (Olle Lundberg, Johan Fritzell, Mikael Rostila, Monica Aberg Yngwe), and Lund (Juan Merlo, Jan and Kristina Sundquist, Martin Lindstrom), Finland (Pekka Puska, Pekka Martinkainen, Jussi Vahtera, Mika Kivimaki, Tuula Oksanen, Marianna Virtanen, Jaana Halonen, Jan Saarela, Ari Haukkala), France (Goldberg, Basile Chaix), UK (Marmot, Steptoe, Brunner, Wilkinson, Pickett, Gindo Tampubolon), Germany (Siegrist), Japan (Katsunori Kondo, Soshi Takao, Takeo Fujiwara, Naoki Kondo, Jun Aida, Ai Ikeda, Cocoro Shirai), Korea (Juhwan Oh, Myounghee Kim, Soongnang Jang), New Zealand and Australia (Tony Blakely, Philippa Howden Chapman, Anne Kavanagh, Billie Giles-Corti, Lisa Wood), Brazil (Alex Chiavegatto Filho, Naomar Almeida Filho), Canada (Arjumand Siddiqi, Spencer Moore, Roman Pabayo, the late and dearly missed Clyde Hertzman) have not only shaped social epidemiology but have influenced much of what we know about occupational epidemiology at the organizational interface. The MacArthur Foundation Networks have again shaped the field in fundamental ways with networks on aging societies and SES led so magnificently by Jack Rowe and Nancy Adler, respectively. The National Institute of Aging Division of Behavioral and Social Research has unfailingly served to promote social epidemiology on a global scale, and huge thanks go to Richard Suzman for his creative force to see what needed to be done to promote the very best social science. Also, our department at Harvard School of Public Health (whose name has changed no less than 3 times in the last decade, reflecting changes in the development of the field) and the Harvard Center for Population and Development Studies have been supported in a compelling way that enabled a vision of social epidemiology and social determinants of health to thrive. We thank all our colleagues here, but especially Jason Beckfield, Amitabh Chandra, Rohini Pande, Mary Waters, and David Canning, who have challenged us as social epidemiologists to define our ideas and approaches with rigor.

Maria Glymour would like to particularly thank Amitabh Chandra for insightful and critical comments about the policy relevance of epidemiology. She thanks Sandy Jencks for his great kindness, which was usually delivered in the disarming guise of engaging arguments about how little we know, and how we might learn more. Theresa Osypuk, David Rehkopf, Stephen Gilman, Kristen Patton, and Mauricio Avendano have been fantastic colleagues for the past several years. Mauricio co-authored several chapters in this volume, but that understates his full intellectual contribution. Maria is especially grateful to Lisa Berkman, Ichiro Kawachi, and Laura Kubzansky, who have been outstanding mentors, colleagues, and friends.

This book would not have been possible without the generous support from so many foundations, notably the Robert Wood Johnson Foundation and the MacArthur Foundation, and from the National Institutes of Health, notably the National Institute on Aging (NIA) and the National Institute of Child Health and Development (NICHD), who have the vision to transcend specific diseases and to orient much of their agenda to important social and economic conditions that shape the lives of people across the entire lifecourse. These institutions have lent their long-term support to efforts to develop the field of social epidemiology. I, LFB, am also grateful for the support that Jim Smith has provided in the form of both intellectual engagement and space and time at RAND during my sabbatical to write this book. It would not have been written without that special time. We would also like to thank the outstanding editorial support from Chad Zimmerman at Oxford University Press.

Of course no volume on social determinants of health could be complete without thanking our closest family and friends. I, LFB, am still amazed that my now adult children who work in very different fields from me understand the work that we are doing in social epidemiology. I thank Andrei and Alex Pogany for all their love and encouragement over the years. Ichiro's hero is Catherine. Maria thanks Stephen Brennan for his patient good humor through many long evenings, and her wonderful family for their love and clarity.

CONTRIBUTORS

Jessica Allen
University College London

Mauricio Avendano
The London School of Economics and Political Science

Lisa F. Berkman
Harvard Center for Population and Development Studies and Harvard School of Public Health

Kirsten Davison
Harvard School of Public Health

Karen Emmons
Kaiser Foundation Research Institute

Thomas A. Glass
Johns Hopkins Bloomberg School of Public Health

M. Maria Glymour
University of California, San Francisco, School of Medicine

Ichiro Kawachi
Harvard School of Public Health

Amii M. Kress
Johns Hopkins Bloomberg School of Public Health

Nancy Krieger
Harvard School of Public Health

Aditi Krishna
Harvard School of Public Health

Laura D. Kubzansky
Harvard School of Public Health

Michael Marmot
University College London

Cassandra Okechukwu
Harvard School of Public Health

Teresa E. Seeman
University of California, Los Angeles Schools of Medicine and Public Health

S. V. Subramanian
Harvard School of Public Health

Töres Theorell
Karolinska Institute

Ashley Winning
Harvard School of Public Health

SOCIAL EPIDEMIOLOGY

CHAPTER 1

A HISTORICAL FRAMEWORK FOR SOCIAL EPIDEMIOLOGY

Social Determinants of Population Health

Lisa F. Berkman and Ichiro Kawachi

Epidemiology is the study of the distribution and determinants of states of health in populations (1). Ever since John Graunt (2) counted deaths in county parishes in England in the seventeenth century, social variations in morbidity and mortality have been observed. Early studies often centered on the ill effects of poverty, poor housing conditions, and work environments. By the nineteenth century, physicians such as Villermé (3) and Virchow (4) refined observations identifying social class and work conditions as crucial determinants of health and disease (5, 6). Chadwick, also a mid-nineteenth-century British public health leader, described the risky physical environment of the poor (7). Durkheim wrote eloquently about another profound social experience, that of social integration and how it was related to patterns of mortality, especially suicide (8). So, in many ways, the idea that social conditions influence health is not new. Social epidemiology, however, is a relatively new area of epidemiology and one that has flourished over the last several decades. In fact, since the first edition of this book the field has grown exponentially (9, 10).

As the public health movement developed in the United States and Great Britain in the nineteenth and early twentieth centuries, attention was drawn to the increased risk of disease among the poor (11, 12). Efforts to improve their physical environments (e.g., housing, noxious work environments, and water supply), sanitation, nutrition, and access to immunization were the primary focus of public health professionals. With broad improvements in the physical environment in the United States, Great Britain, and much of northern Europe, countrywide increases in life expectancy occurred. Based on this observation, many scientists forecast large-scale reductions of social disparities in health (13). Perhaps no other phenomenon has augured the need for the perspective of social epidemiology as clearly, however, as the continued maintenance and recent growth of social inequalities in health in many countries. Thus, while diseases have come and

gone, some infectious diseases have been eradicated, others have emerged, and a host of noninfectious diseases have dominated the profile of causes of death and disability, social inequalities in health remain. These persistent patterns call for an epidemiologic approach to understanding disease etiology that incorporates social experiences as more direct or fundamental causes of disease and disability (14) than is the customary view. Social epidemiology, then, is that branch of epidemiology concerned with the way that social structures, institutions, and relationships influence health. Social epidemiologists are concerned with the ways that societies are organized to produce or impede the development and maintenance of good health. While, as others have noted (15), we share many approaches with other scientists, especially social and behavioral scientists (sociologists, economists, and psychologists), we are differentiated from those areas by a commitment to studying health and ultimately to improving population health with concern for both the overall levels of population health and the distribution of health within populations.

Fortunately, many forces have converged to permit the development of this field. These include: (1) an integration of understanding how social experiences influence physiologic stress responses (16); (2) more nuanced understanding of the distribution of population health growing from Rose's original paradigms; and (3) the evaluation of the health impacts of social and economic policies. When these three approaches are fully integrated with lifecourse issues concerning the accurate identification of etiologic periods and multilevel analytical approaches, they illustrate how our social world powerfully shapes patterns and distributions of health.

THE DIRECT PHYSIOLOGIC INFLUENCES OF SOCIAL EXPERIENCE: BIOLOGICAL EMBEDDING

Among the most critical developments in social epidemiology has been work on stress and physiologic responses to stressful experiences. Building on the fundamental work by Cannon (17) and Selye and Wolff (18), health psychologists, neuroendocrinologists, and physiologists have made it clear that stressful conditions may exact a direct toll on the body, offering powerful biological models that link external stressors to physiologic responses capable of influencing disease development and prognosis. Work on psychophysiology, psychoneuroimmunology, and most recently on allostatic load has helped trace biologic pathways as well as specific behaviors and exposures to noxious agents that link social conditions to important health outcomes (19–22). In the late 1990s, work on the biology of disadvantage took an important turn with a MacArthur Foundation Network on Socioeconomic Status and Health. One of the major accomplishments of this network was related to its work on understanding the biology of stress as it related to social and economic conditions (16). The refinement of concepts related to allostatic load and the role of the hypothalamic-pituitary-adrenal axis, developed in animal models by Bruce McEwen and in epidemiologic work by McEwen and Seeman, emerged from research in the Network. Later work on telomeres, identified initially by Elizabeth Blackburn but furthered in work in social psychology and epidemiology by Epel, was a hallmark of the Network. Both pathways build on the idea that there is a cascade of effects related to early life disadvantage and accumulating

throughout life that dysregulate a number of biological systems leading to the onset and progression of a number of diseases (16).

ROSE'S PARADIGM AND THE DISTRIBUTION OF POPULATION HEALTH

A second theoretical development in understanding the distribution of risk in populations further enhances our ability to launch a solid investigation of social factors and health. In 1992, Geoffrey Rose (6), an eminent epidemiologist, wrote a small book on the strategy of preventive medicine. In this landmark work, small only in size, Rose pointed out that rarely are either risk factors or disease binary in nature. In most cases, risks are distributed along a continuum and small shifts in the distribution of risk throughout a population can make large differences in the health status of that population. Furthermore, understanding the dynamics of why some populations have certain distributions leads to very different etiologic questions than asking why some individuals are in the tails of the distribution. Pursuing this population-based strategy, rather than a high-risk strategy, leads to framing very different questions and utilizing very different preventive approaches. The population strategy is of central importance to social epidemiology and it has been traditionally the mainstay of public health.

EVALUATION OF HEALTH IMPACTS OF PUBLIC POLICIES

Over the last decade economists have started to evaluate the impact of policies—often in the United States at a state level and in cross-country comparisons mostly among Organisation for Economic Co-operation and Development (OECD) countries in terms of health consequences. Since many of the tools (instrumental variable analyses, difference-in-difference models) were popularly used by economists to evaluate policy outcomes, it was a natural outgrowth of these methods to integrate health outcomes into their work. More recently, social epidemiologists have successfully incorporated many of the econometric methods into their work to understand the health consequences of specific policies. Our book now reflects a serious commitment to incorporating both methods and substantive evaluations of social and economic policies into social epidemiology. This is reflected in a number of new chapters on policies (see Chapters 6 and 12). In addition, we have a new chapter by Kawachi on behavioral economics (see Chapter 13) and a chapter on implementation of social determinants of health by Marmot (see Chapter 15).

BACKGROUND

The fields of physiology and psychosomatic, social, and preventive medicine as well as medical sociology and health psychology have all made important contributions to the development of social epidemiology (see 11 for an excellent discussion of preventive medicine in the United States). But the seeds of social epidemiology have also grown from within epidemiology itself. In

1950, Alfred Yankauer, in the American Sociological Review wrote an article on infant mortality in which the subhead was "an inquiry into social epidemiology" (23). Nancy Krieger credits this with being the first mention of the term "social epidemiology" (24). E. Gartley Jaco further mentions the term "social epidemiology" in two volumes published in 1958 and 1960 (25, 26). The field was born, and by the late 1960s and 1970s epidemiologists such as Leo Reeder, John Cassel, Mervyn Susser, S. Leonard Syme, Saxon Graham, Lawrence Hinkle, Al Tyroler, and Sherman James started to develop a distinct area of investigation in epidemiology centered on the health impact of social conditions, particularly cultural change, social status and status inconsistency, and life transitions. Their work drew heavily on that of epidemiologists who worked earlier in the century such as Goldberger and Sydenstricker (27), who investigated the etiology of pellagra, and Wade Hampton Frost, whose work on tuberculosis was seminal (28). They also drew deeply from medical sociology (29) and the work of psychiatric epidemiologists (30–33). Syme (34) explained that investigations of the "social etiology of disease attempted to systematically examine variations in the incidence of particular diseases among people differentially located in the social structure and attempt[ed] to explore the ways in which their position in the social structure tended to make them more vulnerable, or less, to particular disease."

In a seminal article, Saxon Graham (35) discussed the social epidemiology of selected chronic illnesses. While never giving an explicit definition of social epidemiology, he suggested that a union of sociology with the medical sciences would produce a new and more successful epidemiology. Graham went on to say that achieving a coherent and complete theory of disease causation would require obtaining social and biological data that are consistent with each other with regard to a specific disease (35). Thus, he argued, one must understand how membership in a social group relates to behavior patterns, to exposure to "vehicles" for transmitting agents, to direct tissue changes, and finally to disease. Graham aimed to identify specific social circumstances that led to a chain of events in which specific behaviors were linked to specific diseases. His classic example involved Percival Pott's analysis of scrotal cancer in chimney sweeps. Parallel to his analysis of Pott's studies, much of his early work dealt with smoking and dietary and sexual behaviors that were associated with different social groups and thus more proximally linked to specific diseases. In seeking to understand the large-scale social patterning of disease in terms of individual behaviors of group members, Graham's great contribution to epidemiology was his ability to incorporate this multilevel thinking into the field.

Almost a decade later, in the mid-1970s, two epidemiologists, John Cassel and Mervyn Susser, more explicitly tackled the methodologic controversies and paradigm shifts inherent in incorporating a deeper understanding of the social influences of disease into epidemiologic thinking. Armed with evidence from the previous decade, John Cassel (36) in the fourth Wade Hampton Frost Lecture to the American Public Health Association stated that "the question facing epidemiologic inquiry is, are there categories or classes of environmental factors that are capable of changing human resistance in important ways and making subsets of people more or less susceptible to ubiquitous agents in our environment." In his classic paper "The Contribution of the Social Environment to Host Resistance," he argued that environmental conditions capable of "producing profound effects on host susceptibility" involve the presence of other members of the same species, or more generally, certain aspects of the social environment (36).

Building on the work of Hinkle (37) and stress researchers such as Cannon (17), Dubos (38), and Selye and Wolff (18), Cassel posited that at least one of the properties of

stressful situations might be that the actor is not receiving adequate evidence that his actions are leading to anticipated consequences. Today we might cite situations of powerlessness brought on by social disorganization, migration, discrimination, poverty, and low support at work as prime examples of this situation. Cassel also outlined a series of protective factors that might buffer the individual from the deleterious consequences of stressful situations. The property common to these processes is "the strength of the social supports provided by the primary groups of most importance to the individual" (36). Thus, consolidating the findings gathered by epidemiologists doing empirical work on status and status incongruity (34, 37), rapid social change and disorganization (39, 40), acculturation and migration (41), and social support and family ties (42, 43), Cassel laid out an intellectual agenda for social epidemiology that provided the groundwork for decades to come.

In a provocative series of articles, Mervyn Susser wrote that epidemiology must broaden its base and move beyond its focus on individual-level risk factors and "black box epidemiology" to a new "multilevel ecoepidemiology" (44–48). The foundations for much of this framework can be seen in his 1973 book, *Causal Thinking in the Health Sciences: Concepts and Strategies in Epidemiology*. In the introduction to that volume, Susser stated that epidemiology shares the study of populations, in a general way, with other population sciences such as sociology, human biology, and population genetics. In affirming common methodologic and conceptual ground with other sciences involved in the study of society, he explained that "states of health do not exist in a vacuum apart from people. People form societies and any study of the attributes of people is also a study of the manifestations of the form, the structure and the processes of social forces" (1). In other chapters, Susser discussed how agent, host, and environment models, the most basic organizing principles of epidemiology, could be framed as an ecological system with different levels of organization.

Susser again emphasizes that epidemiology is, in essence, ecological, since the biology of organisms is determined in a multilevel, interactive environment. Identifying risks at the individual level, even multiple risks, does not sufficiently explain interactions and pathways at that level, nor does it incorporate the social forces that influence risks to individuals.

Most recently, Nancy Krieger has developed several distinct theories and frameworks that explicitly frame how and why social experiences become biologically embedded, including a conceptual model drawing on an ecosocial theory. The central question for ecosocial theory is: "who and what is responsible for population patterns of health, disease, and well-being, as manifested in present, past and changing social inequalities in health" (24). Core concepts of the theory include embodiment, or the ways in which we incorporate, biologically, the material and the social world in which we live from in utero until death; pathways to embodiment; the cumulative interplay between exposure, susceptibility, and resistance; and accountability and agency.

GUIDING CONCEPTS IN SOCIAL EPIDEMIOLOGY

We define social epidemiology as the branch of epidemiology that studies the social distribution and social determinants of states of health. Defining the field in this way implies that we aim to identify socioenvironmental exposures that may be related to a broad range of physical

and mental health outcomes. Our orientation is similar to other subdisciplines of epidemiology focused on exposures (e.g., environmental or nutritional epidemiology) rather than those areas devoted to the investigation of specific diseases (e.g., cardiovascular, cancer, or psychiatric epidemiology). We focus on specific social phenomena such as socioeconomic stratification, social networks, discrimination, workplace organization, and public policies rather than on specific disease outcomes. While future studies may reveal that some diseases are more heavily influenced by social experiences than others, we suspect that the vast majority of diseases and other health outcomes such as functional status, disability, and well-being are affected by the social world surrounding us all.

Like environmental and nutritional epidemiology, social epidemiology must integrate phenomena at the margins of what is defined as its domain. For instance, psychological states, behaviors, and aspects of the physical or built environment are influenced by social environments and vice versa. Borders at the periphery of any field—and social epidemiology is no exception—are bound to be fuzzy. We make no attempt to draw clean lines encircling the field. Because it is important for social epidemiologists to consider related areas, in this volume we have included chapters on psychological states and behaviors that are closely related to the social experiences that are our primary concern. If we err on the side of blurring boundaries, we must balance that with precision in defining explicit testable hypotheses in our work. Without hypotheses that can be clearly supported or refuted, without having a clear understanding of temporal sequencing or biological plausibility, and without articulated theories and specific concepts to guide empirical investigation, we will not be able to make progress.

The rest of this chapter outlines several concepts that are important to the field of social epidemiology. These concepts are not offered as universals to be uncritically accepted but rather as useful and sometimes challenging guides that transcend the study of any single exposure.

A POPULATION PERSPECTIVE

Individuals are embedded in societies and populations. The crucial insight provided by Rose's (6) population perspective is that an individual's risk of illness cannot be considered in isolation from the disease risk of the population to which she belongs. Thus, a person living in Finland is more likely to die prematurely of a heart attack compared with someone living in Japan, not just because any particular Finnish individual happens to have a high level of cholesterol but because the population distribution of cholesterol levels in Finnish society *as a whole* is shifted to the right of the Japanese distribution. The level of cholesterol that might be considered "normal" in Finnish society would be grossly abnormal and a cause for alarm in Japan. Moreover, we know from detailed studies of migrants that the basis for these population differences is not genetic (41). For instance, Japanese immigrants to America take on the coronary risk profiles of their adopted country.

Although Rose's initial examples involved the examination of risk factors for heart disease, we now recognize that his insight has broad applicability to a swath of public health problems, ranging from aggression and violence, to mental health, to the effects of poverty and material deprivation on health. Fundamentally, Rose's insight harks back to Durkheim's discovery about suicide: that

the rate of suicide in a society is linked to collective social forces. There are myriad reasons why any individual commits suicide, yet such individuals come and go while the *social* rate of suicide remains predictable.

The crucial implication of Rose's theory for social epidemiology is that we must incorporate the social context into explanations about why some people stay healthy while others get sick. Applying the population perspective into epidemiological research means asking, "Why does *this* population have *this* particular distribution of risk?" in addition to asking, "Why did this particular individual get sick?" Furthermore, as Rose pointed out, the greatest improvements in population health are likely to derive from answering the first question, because the majority of cases of illness arise within the bulk of the population who are outside the tail of high risk. The central feature of early work in this area suggested that population distributions for most risk factors move along a continuum with a normal distribution. Rose said the "distribution of health-related characteristics move up and down as a whole: the frequency of cases can be understood only in the context of the population's characteristics" (6). For instance, in the United States, obesity has been on the rise over the past few decades, and a careful examination supports Rose's proposition that the entire risk distribution has shifted during this time. In fact, the mean has driven the tail.

As evidence has accumulated to test these hypotheses, it has become apparent that there are variations in these patterns. This is important if increasingly more people are in the tails of the distribution or if the distribution itself is not normally distributed. In these situations, further evaluation of Rose's strategies may be critical, and specific high-risk strategies may become more effective. In fact, this rather straightforward empirical question has just been examined. Similarly, if relative risks increase non-linearly or show clear threshold effects, different actions may be appropriate. Depending on the distribution of the risk factor and the pattern of the effect of the risk factor on disease, a population strategy of risk reduction may not always produce the same results. Rose himself acknowledged that under certain circumstances, high-risk strategies may be more effective on a population level than population-wide strategies. Certainly, from a purely hypothetical perspective, if the risk distribution is very skewed, Rose's paradigm may be more problematic. The issue at hand is whether this is actually the case with regard to some risks or if it is a purely hypothetical situation.

Another issue also relates to the shape of the curve. For instance, considering the issue of economic inequality and health, one can imagine two curves: both normally distributed but with very different standard deviations. If there is something harmful about inequality itself, its relative nature, not just the absolute prevalence of poverty, then shifting the curve to the left or giving everyone the same amount of money will do little to improve population health. Reducing the standard deviation or the percentage of people in the tails of the distribution is critical to improving health. In this case, the mean may stay exactly the same, but the tails move toward the center. One could imagine population strategies that might produce a tighter risk distribution around the mean or one could employ specific high-risk strategies to pull in the tails, especially the tail with highest risk. In any case, it is time for a second critical look at the population paradigm developed by Rose and to subject the theories to empirical tests so that optimal strategies to improve population health will be developed.

More recent work suggests that health transitions may occur in which the distribution itself does not move as a whole but rather stretches out, creating increasing differences

between those at one tail of the distribution and those at the other (49). Rose postulated that for most risk factors, as the average level in the population changed, the "dispersion around the average remains rather constant" (6). However, observations in some countries more recently suggest that not all risk distributions may follow this pattern, and that a proportionally greater increase may appear at tails of risk distributions creating more inequalities and nonnormal distributions of risk. The distribution of BMI in low- and middle-income countries may follow this sort of pattern. This pattern of increased weight gain among higher percentiles of the BMI distribution has been noted in a few national surveys of children and adults (49).

With his emphasis on focusing on the distributions of risk in populations, Rose made a clear contribution to the history of social epidemiology. Now the task is to refine his theories based on the growing empirical base of data available across the world and over time. As a further note, Krieger has suggested that we become more reflective of how we define populations and explicitly consider who is included in definitions of populations (50).

THE SOCIAL AND ECONOMIC CONTEXT OF BEHAVIOR: RISKS OF RISKS

Over the last several decades, a huge number of clinical trials have been launched to modify individual behavioral risk factors such as alcohol and tobacco consumption, diet, and physical activity. By and large, the most successful have been those which incorporated elements of social organizational changes into interventions. We now understand that behaviors are not randomly distributed in the population. Rather, they are socially and economically patterned and often cluster with one another. Thus, many people who drink also smoke cigarettes, and those who follow health-promoting dietary practices also tend to be physically active. People who are poor, have low levels of education, or are socially isolated are more likely to engage in a wide range of risk-related behaviors and less likely to engage in health-promoting behaviors (51, 52). This patterned behavioral response has led Link and Phelan (14) to speak of situations that place individuals "at risk of risks."

Understanding why "poor people behave poorly" (53) requires a shift in understanding—specific behaviors once thought of as falling exclusively within the realm of individual choice occur in a social context. The social environment influences behavior by *(1)* shaping norms, *(2)* enforcing patterns of social control (which may be health-promoting or health-damaging), *(3)* providing or not providing environmental opportunities to engage in certain behaviors, and *(4)* reducing or producing stress for which certain behaviors may be an effective coping strategy, at least in the short term. Environments place constraints on individual choice and incentivize particular choices with promises of social, psychological, financial, or physical rewards. Incorporating the social context into behavioral interventions has led to a whole new range of clinical trials that take advantage of communities, schools, and worksites to achieve behavioral change (see 54 and Chapter 10 on health behaviors).

THE CASE OF TOBACCO USE

Over the last decades, in the United States, tobacco consumption has decreased substantially with evident health gains clear in the subgroups of the population where consumption has fallen, and worsening health among those whose consumption increased over the last decades. The decline in tobacco use in the United States is attributed to a number of factors; it is clear that no single intervention is responsible for the decline. What is clear is that individually oriented cessation programs played a minor role in changing the course. Rather, environmental and policy changes including taxation, prohibition of smoking in public settings, and limits on advertising and sales to minors all played a role. This public health/policy attack on consumption highlights the importance of social and economic level interventions rather than aiming for individual behavior changes. The current social patterning of tobacco use simultaneously indicates where vulnerabilities and disadvantage still shape consumption and calls for the continued need to decrease risks in socially disadvantaged groups. A social epidemiological approach calls, above all, for an understanding of the complex social and economic dynamics driving what seems to be, but is not, individual choice.

CONTEXTUAL MULTILEVEL ANALYSIS

The understanding that behavior is conditioned by society yields a more general appreciation of the need for contextual analysis in epidemiology. As Susser (46) noted, "risk factor epidemiology in its pure form exploits neither the depth and precision of micro-levels nor the breadth and compass of macro-levels." Conceptions of how culture, policy, and the environment influence health remain fuzzy and speculative if one analyzes only the independent effects of individual-level risk factors. Ecological analysis, a central part of both epidemiology and sociology early in this century, offered an approach to the study of environments, but it lost a great deal of respectability because of problems related to the ecological fallacy (i.e., drawing individual inferences from grouped data). It was difficult, if not impossible, to rule out reverse causation (that the illness influenced residential relocation) in many studies. In fact, it was this latter problem that plagued many of the early studies on psychiatric disorder and community disorganization.

In the past few years, however, it has become apparent that just as there are ecologic-level exposures in environmental and infectious disease epidemiology, so are there valid ecologic-level exposures related to the social environment that are not adequately captured by investigation at an individual level (55–58). For example, the number of grocery stores or parks, the condition of housing stock, and voter participation may be critical determinants of behaviors, access to care, or illness. These ecologic-level exposures call for innovative methods (59–61). The assessment of exposures at an environmental or community level may lead to an understanding of social determinants of health that is more than the sum of individual-level measures. Although important questions remain about the appropriate level of environmental assessment (e.g., neighborhood, city, state, country), the disentangling of compositional versus contextual effects, and the pathways linking such environmental exposures to individual health outcomes, ecological analyses offer a valuable research tool to epidemiologists. When coupled with individual-level data, they offer the critical advantages available in the form of multilevel analyses.

A DEVELOPMENTAL AND LIFECOURSE PERSPECTIVE

The integration of lifecourse epidemiology into almost all epidemiologic inquiries has vastly improved the field of epidemiology generally and social epidemiology more specifically. The identification of etiologic periods (for both causing illness as well as changing exposures) is central to improving population health. Three main lifecourse trajectories have been well defined. First, developmentalists interested in early development and childhood have focused for decades on the importance of early life exposures in shaping cognition and brain function (62). Over the past two decades, epidemiologists have come to understand the early origins of diseases, often focusing on fetal origins, which evidence suggests shape patterns of metabolic function related to diabetes and other health outcomes (62). Power and Hertzman (63), for example, propose that exposure in early childhood could influence developmental processes—particularly brain development during periods of great plasticity. By molding patterns of response during these "critical stages," early life experiences would then make the individual vulnerable or resistant to various diseases in adulthood (64). Second, many epidemiologists interested in lifecourse issues hypothesize that most adult disease is not likely the result of early childhood or prenatal exposure but rather the result of a lifetime of accumulated exposures (62). Such a trajectory can incorporate early exposures and adult exposures simultaneously because it is the impact of cumulative exposures across the lifecourse that takes a toll at older ages. Early experiences may produce some independent impact on outcomes, but that is not the central issue in this model. In this model, the etiologic period is long and covers decades of an individual's life, starting either in early childhood or in adulthood. Third, in a social trajectory model of health and disease, early life exposures impact adult exposures, which in turn directly influence disease risk. In these social trajectory models, the causal pathways indicate that early life exposures do not directly affect adult health. They influence adult social conditions, which, in turn, affect adult health. In this case, intervention in adulthood can completely offset harm incurred in childhood.

The lifecourse model of trajectories of risk is well articulated theoretically but much harder to test empirically. Yet there is intriguing evidence that such perspectives may yield valuable insights. In fact, social epidemiologists working in the 1960s and 1970s implicitly adopted a lifecourse perspective in testing theories about status incongruity in which the stressful experiences being studied resulted from having grown up in one situation or as a member of one status group and then having shifted to either a higher or lower status. (See 33 for an excellent discussion of this.)

These three models lay out a framework within which to examine the ways in which social and economic conditions may influence health over the lifecourse. Our aim here is not to conclude that there is strong evidence to support one or another of them, nor in fact to advocate an overly deterministic, developmental model of disease causation at all, but rather to suggest that this perspective provides a lens through which to examine how social factors may influence adult health.

RESISTANCE AND SUSCEPTIBILITY TO DISEASE: ACCELERATED AGING

Wade Hampton Frost (65) noted that at the turn of the twentieth century there was nothing that changed "nonspecific resistance to disease" as much as poverty and poor living conditions. In referring to this altered resistance, Frost suggested that it was not just increased risk of exposure among the poor that produced a high prevalence of tuberculosis: It was something about their inability to fight off the disease—their increased susceptibility to disease *once* exposed—that contributed to higher rates of disease in poor populations.

Cassel, Syme, and Berkman (36, 66, 67) built on this idea when they observed that many social conditions were linked to a very broad array of diseases and disabilities. They speculated that social factors influence disease processes by creating a vulnerability or susceptibility to disease in general rather than to any specific disorder. According to the general susceptibility hypothesis, whether individuals developed one disease or another depended on their behavioral or environmental exposures as well as their biological or genetic makeup. But whether they became ill or died at earlier ages or whether specific socially defined groups had greater rates of disease depended on socially stressful conditions.

As originally proposed, the concept of general susceptibility or psychosocial "host resistance" was a powerful and intuitively appealing metaphor but not well grounded biologically. It was not until research in social epidemiology became more integrated with research in neuroscience and psychoneuroimmunology that clear biological mechanisms were defined, at least as potential pathways leading from stressful social experiences to poor health. Neuroendocrinologists had identified classic stress mediators such as hypothalamic-pituitary-adrenal axis-related deregulation as well as inflammatory markers, telomeres, and other less well understood mediators such as dehydroepiandrosterone (DHEA), prolactin, and growth hormone, and they knew that these affected multiple physiologic systems. By linking evidence from both fields, researchers showed that some stressful experiences activate multiple hormones and inflammatory processes and thus might not only affect multiple systems but could also produce wide-ranging end-organ damage. Furthermore, recent advances in understanding variable patterns of neuroendocrine response with age suggest that the cumulative effects of stress, or even stressful experiences that have taken place during development, may alter neuroendocrine-mediated biological pathways and lead to a variety of disorders from cardiovascular disease to cancer and infectious disease (22, 68, 69).

These developments in aging research suggest new ways in which stressful experiences may be conceptualized as accelerating the rate at which we age or changing the aging process itself (70). This conceptual shift relates well to earlier notions of general susceptibility but refines the links in substantial ways. Recent work on telomeres, stress reactivity, allostatic load, and inflammatory processes more generally all point to the dysregulating role that stress from social conditions as well as the linked behavioral pathways play in aging processes. For instance, age is associated with telomere length, and telomeres have been linked to mortality and morbidity rates (71–75). While the social and economic determinants of telomere length are not yet clear, there is growing evidence that shorter telomeres are associated with chronic stress (76), as well as lower socioeconomic position,

less educational attainment, and unemployment (77–81). It is of great interest to identify the links between social circumstances and biological aging, as this will provide clues about the ubiquitous associations between social conditions and such a large array of diseases and disabling conditions. If such conditions lean on the physiological systems linked to dysregulation and aging processes, accelerated aging may be key to understanding social determinants of health.

CONCLUSION

In recent decades, the discipline of epidemiology has witnessed the birth of multiple subspecialties such as environmental, nutritional, clinical, reproductive, and most recently, genetic epidemiology (82). The central question of social epidemiology—how social conditions give rise to patterns of health and disease in individuals and populations—has been around since the dawn of public health. But the rediscovery of this question through the lens of epidemiology is a relatively recent phenomenon. As demonstrated in the contributions to this volume, social epidemiologists are now applying concepts and methods imported from a variety of disciplines ranging from sociology, psychology, political science, economics, demography, and biology. The multidisciplinary nature of the venture makes the research both new and suited to tackle the problems at hand. Social epidemiology has already yielded many important findings during the relatively brief period of its existence, yet important discoveries remain to be made. By sharpening the tools we have to capture the powerful social forces experienced by individuals and communities, as well as by strengthening our methods of inquiry, we may look forward to further decades of insight into how society shapes the health of people. With rigorous attention to issues related to the social context, biological mechanisms, the timing and accumulation of risk, and optimal moments for intervention, we can hope to identify the ways in which the structure of society influences the public's health.

REFERENCES

1. Susser M. Causal thinking in the health sciences: concepts and strategies of epidemiology. New York: Oxford University Press; 1973.
2. Graunt J. Natural and political observations mentioned in a following index and made upon the bills of mortality. London (1939): Reprinted Johns Hopkins University Press, Baltimore; 1662.
3. Villermé LR. De la mortalité dans divers quartiers de la ville de Paris. Annales d'hygiene publique. 1830;3:294–341.
4. Virchow R. Report on the typhus epidemic in Upper Silesia. In: Rather JJ, editor. Rudolph Virchow: collected essays on public health and epidemiology. Canton, MA: Science History; 1848. pp. 205–20.
5. Rosen G. The evolution of social medicine. In: Freeman HE, Levine S, Reeder LG, editors. Handbook of medical sociology. Englewood Cliffs, NJ: Prentice Hall; 1963. p. 61.
6. Rose GA. The strategy of preventive medicine. New York: Oxford University Press; 1992.

7. Chadwick E. Report on the sanitary condition of the labouring population of Great Britain. London: Poor Law Commission; 1842.
8. Durkheim E. Suicide: a study in sociology. Glencoe, IL: Free Press; 1897.
9. Galobardes B, Lynch JW, Smith GD. Is the association between childhood socioeconomic circumstances and cause-specific mortality established? Update of a systematic review. J Epidemiol Community Health. 2008;62(5):387–90.
10. Holt-Lunstad J, Smith TB, Layton JB. Social relationships and mortality risk: a meta-analytic review. PLoS Med. 2010;7(7):e1000316.
11. Rosen G. Preventive medicine in the United States 1900–1975: trends and interpretation. New York: Science History; 1975.
12. Duffy J. The sanitarians: a history of American public health. Chicago: University of Illinois Press; 1990.
13. Kadushin C. Social class and the experience of ill health. Sociol Inq. 1964;35:67–80.
14. Link BG, Phelan J. Social conditions as fundamental causes of disease. J Health Soc Behav. 1995;Spec No:80–94.
15. Brandt AM, Gardner M. Antagonism and accommodation: interpreting the relationship between public health and medicine in the United States during the 20th century. Am J Public Health. 2000;90(5):707–15.
16. Adler NE, Stewart J. Health disparities across the lifespan: meaning, methods, and mechanisms. Ann N Y Acad Sci. 2010;1186(1):5–23.
17. Cannon WB. Stresses and strains of homeostasis. Am J Med Sci. 1935;189:1–14.
18. Selye H, Wolff HG. The concept of "stress" in the biological and social sciences. Sci Med Man. 1973;1:31–48.
19. Cohen S. Psychosocial models of the role of social support in the etiology of physical disease. Health Psychol. 1988;7(3):269–97.
20. Kiecolt-Glaser JK, Glaser R, Cacioppo JT, MacCallum RC, Snydersmith M, Kim C, et al. Marital conflict in older adults: endocrinological and immunological correlates. Psychosom Med. 1997;59(4):339–49.
21. Kiecolt-Glaser JK, Glaser R, Gravenstein S, Malarkey WB, Sheridan J. Chronic stress alters the immune response to influenza virus vaccine in older adults. PNAS. 1996;93(7):3043–7.
22. McEwen BS. Protective and damaging effects of stress mediators. N Engl J Med. 1998;338(3):171–9.
23. Yankauer A. The relationship of fetal and infant mortality to residential segregation: an inquiry into social epidemiology. Am Sociol Rev. 1950;15(5):644–8.
24. Krieger N. Theories for social epidemiology in the 21st century: an ecosocial perspective. Int J Epidemiol. 2001;30(4):668–77.
25. Jaco EG. Introduction: medicine and behavioral science. In: Physicians and illness: sourcebook in behavioral science and medicine. Glencoe, IL: The Free Press; 1958. pp. 3–8.
26. Jaco EG. The social epidemiology of mental disorders: a psychiatric survey of Texas. New York: Russell Sage Foundation; 1960.
27. Goldberger J, Wheeler E, Sydenstricker E, King WI. A study of endemic pellagra in some cotton-mill villages of South Carolina. Hygienic Laboratory Bulletin. 1929;153:1–66.
28. Maxcy KF, editor. Papers of Wade Hampton Frost. New York: Commonwealth Fund; 1941.
29. Freeman HE, Levine S, Reeder LG. Handbook of medical sociology. Englewood Cliffs, NJ: Prentice-Hall; 1963.
30. Faris R, Dunham HW. Mental disorders in urban areas. University of Chicago Press; 1939.
31. Hollingshead AB, Redlich FC. Social class and mental illness. New York: John Wiley; 1958.
32. Leighton AH. My name is legion. New York: Basic Books; 1959.
33. Srole L. Mental health in the metropolis. New York: McGraw-Hill; 1962.

34. Syme SL, Hyman MD, Enterline P. Cultural mobility and the occurrence of coronary heart disease. J Health Human Behav. 1965;6:178–89.

35. Graham S. Social factors in relation to chronic illness. In: Freeman H, Levine S, Reeder LG, editors. Handbook of medical sociology. New Jersey: Prentice-Hall; 1963.

36. Cassel J. The contribution of the social environment to host resistance. Am J Epidemiol. 1976;104(2):107–23.

37. Hinkle LE. The concept of "stress" in the biological and social sciences. Sci Med Man. 1973;1:31–48.

38. Dubos R. Man adapting. New Haven: Yale University Press; 1965.

39. Cassel J, Tyroler H. Epidemiological studies of culture change: I. Health status and recency of industrialization. Arch Environ Health. 1961;3:25–33.

40. James SA, Kleinbaum DG. Socio-ecologic stress and hypertension related mortality rates in North Carolina. Am J Public Health. 1976;66:354–8.

41. Marmot M, Syme SL. Acculturation and coronary heart disease in Japanese-Americans. Am J Epidemiol. 1976;104:224–47.

42. Nuckolls K, Cassel J, Kaplan B. Psychosocial assets, life crisis and the prognosis of pregnancy. Am J Epidemiol. 1972;95:431–41.

43. Pless IB, Saterwaite B. Chronic illness in childhood: selection, activities and evaluation of non-professional family counselors. Clin Pediatr. 1972;11:403–10.

44. Susser M, Susser E. Choosing a future for epidemiology: I. Eras and paradigms. Am J Public Health. 1996;86(5):668–73.

45. Susser M, Susser E. Choosing a future for epidemiology: II. From black box to Chinese boxes and eco-epidemiology. Am J Public Health. 1996;86(5):674–7.

46. Susser M. Does risk factor epidemiology put epidemiology at risk? Peering into the future. J Epidemiol Community Health. 1998;52(10):608–11.

47. Susser M. The logic in ecological: I. The logic of analysis. Am J Public Health. 1994;84(5):825–9.

48. Susser M. The logic in ecological: II. The logic of design. Am J Public Health. 1994;84(5):830–5.

49. Razak F, Corsi DJ, Subramanian SV. Change in the body mass index distribution for women: analysis of surveys from 37 low- and middle-income countries. PLoS Med. 2013;10(1):e1001367.

50. Krieger N. Who and what is a "population?" Historical debates, current controversies, and implications for understanding "population health" and rectifying health inequities. Milbank Q. 2012;90(4):634–81.

51. Matthews K, Kelsey S, Meilahn E. Educational attainment and behavioral and biologic risk factors for coronary heart disease in middle-aged women. Am J Epidemiol. 1989;129:1132–44.

52. Adler NE, Boyce T, Chesney MA, Cohen S, Folkman S, Kahn RL, et al. Socioeconomic status and health: the challenge of the gradient. Am Psychol. 1994;49(1):15–24.

53. Lynch JW, Kaplan GA, Salonen JT. Why do poor people behave poorly? Variation in adult health behaviours and psychosocial characteristics by stages of the socioeconomic lifecourse. Soc Sci Med. 1997;44(6):809–19.

54. Sorensen G, Emmons K, Hunt MK, Johnston D. Implications of the results of community intervention trials. Annu Rev Public Health. 1998;19:379–416.

55. Macintyre S, Maciver S, Sooman A. Area, class and health: should we be focusing on places or people? J Soc Pol. 2009;22(2):213.

56. Kawachi I, Kennedy BP. Health and social cohesion: why care about income inequality? BMJ. 1997;314(7086):1037–40.

57. Kawachi I, Kennedy BP, Lochner K, Prothrow-Stith D. Social capital, income inequality, and mortality. Am J Public Health. 1997;87(9):1491–8.

58. Kaplan GA. People and places: contrasting perspectives on the association between social class and health. Int J Health Services. 1996;26(3):507–19.

59. Jones K, Moon G. Medical geography: taking space seriously. Prog Hum Geog. 1993;17(4):515–24.

60. Diez Roux AV, Nieto FJ, Muntaner C, Tyroler HA, Comstock GW, Shahar E, et al. Neighborhood environments and coronary heart disease: a multilevel analysis. Am J Epidemiol. 1997;146(1):48–63.

61. Subramanian SV, Jones K, Kaddour A, Krieger N. Revisiting Robinson: The perils of individualistic and ecologic fallacy. Int J Epidemiol. 2009;38(2):342–60.

62. Berkman LF. Social epidemiology: Social determinants of health in the United States: are we losing ground? Annu Rev Public Health. 2009;30(1):27–41.

63. Power C, Hertzman C. Social and biological pathways linking early life and adult disease. British Medical Bulletin. 1997;53(1):210–21.

64. Barker D. Fetal and infant origins of adult disease. BMJ. 1990;301(6761):1111.

65. Frost WH. How much control of tuberculosis? Am J Public Health. 1937;27:759–66.

66. Berkman LF, Syme SL. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. Am J Epidemiol. 1979;109(2):186–204.

67. Syme SL, Berkman LF. Social class, susceptibility and sickness. Am J Epidemiol. 1976;104(1):1–8.

68. Meaney MJ, Mitchell JB, Aitken DH, Bhatnagar S, Bodnoff SR, Iny LJ, et al. The effects of neonatal handling on the development of the adrenocortical response to stress: implications for neuropathology and cognitive deficits in later life. Psychoneuroendocrinology. 1991;16(1–3):85–103.

69. Sapolsky RM. Why stress is bad for your brain. Science. 1996;273(5276):749–50.

70. Berkman LF. The changing and heterogeneous nature of aging and longevity: A social and biomedical perspective. Annu Rev Gerontol Geriatr. 1988;8:37–68.

71. Lee W-W, Nam K-H, Terao K, Yoshikawa Y. Age-related telomere length dynamics in peripheral blood mononuclear cells of healthy cynomolgus monkeys measured by Flow FISH. Immunology. 2002;105(4):458–65.

72. Brouilette SW, Moore JS, McMahon AD, Thompson JR, Ford I, Shepherd J, et al. Telomere length, risk of coronary heart disease, and statin treatment in the West of Scotland Primary Prevention Study: a nested case-control study. Lancet. 2007;369(9556):107–14.

73. Bakaysa SL, Mucci LA, Slagboom PE, Boomsma DI, McClearn GE, Johansson B, et al. Telomere length predicts survival independent of genetic influences. Aging Cell. 2007;6(6):769–74.

74. Cawthon RM, Smith KR, O'Brien E, Sivatchenko A, Kerber RA. Association between telomere length in blood and mortality in people aged 60 years or older. Lancet. 2003;361(9355):393–5.

75. Honig LS, Schupf N, Lee JH, Tang MX, Mayeux R. Shorter telomeres are associated with mortality in those with APOE epsilon4 and dementia. Ann Neurol. 2006;60(2):181–7.

76. Epel ES, Merkin SS, Cawthon R, Blackburn EH, Adler NE, Pletcher MJ, et al. The rate of leukocyte telomere shortening predicts mortality from cardiovascular disease in elderly men. Aging. 2009;1(1):81–8.

77. Batty GD, Wang Y, Brouilette SW, Shiels P, Packard C, Moore J, et al. Socioeconomic status and telomere length: the West of Scotland Coronary Prevention Study. J Epidemiol Community Health. 2009;63(10):839–41.

78. Cherkas LF, Aviv A, Valdes AM, Hunkin JL, Gardner JP, Surdulescu GL, et al. The effects of social status on biological aging as measured by white-blood-cell telomere length. Aging Cell. 2006;5(5):361–5.

79. Steptoe A, Hamer M, Butcher L, Lin J, Brydon L, Kivimäki M, et al. Educational attainment but not measures of current socioeconomic circumstances are associated with leukocyte telomere length in healthy older men and women. Brain Behav Immun. 2011;25(7):1292–8.

80. Needham BL, Adler N, Gregorich S, Rehkopf D, Lin J, Blackburn EH, et al. Socioeconomic status, health behavior, and leukocyte telomere length in the National Health and Nutrition Examination Survey, 1999–2002. Soc Sci Med. 2013;85:1–8.

81. Surtees PG, Wainwright NWJ, Pooley KA, Luben RN, Khaw KT, Easton DF, et al. Educational attainment and mean leukocyte telomere length in women in the European Prospective Investigation into Cancer (EPIC)-Norfolk population study. Brain Behav Immun. 2012;26(3):414–8.

82. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.

CHAPTER 2

SOCIOECONOMIC STATUS AND HEALTH

M. Maria Glymour, Mauricio Avendano, and Ichiro Kawachi

INTRODUCTION

The importance of socioeconomic conditions for health has been recognized for centuries (1–3). No matter whether we are talking about the mill towns of Victorian England (4), the sweatshops of New York in the Gilded Age (5), or the slums of Mumbai in contemporary India (6), the poor in every society tend to have worse health and shorter lives than the wealthy. Even comparing societies and time periods in which the leading causes of death are completely different, the socioeconomic pattern in early mortality is almost always the same: the poor die first. Socioeconomic inequalities[1] in health are apparent at almost every stage in the lifecourse, from birth (neonatal outcomes, infant mortality) to working age (e.g., cardiovascular disease, accidents) and old age (functional disability). Lower socioeconomic status (SES) is correlated with increased risks of nearly every major cause of premature mortality (7). Moreover, socioeconomic inequalities in health status are not just a threshold effect of poverty; there is a "gradient" in health across the SES hierarchy such that the higher the level of household income, wealth, education, or occupational ranking, the lower the risks of morbidity and mortality. The gradient of health is observed almost throughout the range of socioeconomic status, so that the middle class have better health than the poor, and the wealthy have better health than the middle class (2).

Socioeconomic status is typically characterized along three dimensions: education, employment, and money. Major health inequalities prevail along all three dimensions. The lower an individual's position in the occupational hierarchy of a workplace, the worse their health status; and the lower someone's educational attainment, the lower their health

1 We primarily use the term "inequalities" in this chapter, although differences in health between groups are more commonly described as "disparities" in the United States.

achievement. These stylized facts have led Link and Phelan to propose that SES is a "fundamental" determinant of health in society; that is, no matter what health threats are relevant to a population at any given time, higher SES individuals and groups are more likely to possess—and to use—more resources (such as knowledge, money, prestige, and power) to protect their health (8).

In this chapter, we summarize the theory and evidence on two dimensions of socioeconomic status—schooling and income—as potential determinants of health. From a public health standpoint, the correlation between SES and health is potentially important for two reasons. The simplest application is to use SES as a marker to identify individuals in need of extra services, for example, modified routines of medical care. The more interesting application is to understand how to intervene on some aspect of SES to improve health. This latter application has promise only if SES is causally related to health, that is, if interventions on SES would in fact influence health.

We argue that although schooling and income are both important drivers of health, not every instance of their correlation with health outcomes is likely to be causal. Some part of the relation between SES and health is likely to reflect reverse causation and confounding by third variables. Furthermore, precisely because SES influences so many pathways leading to health, different approaches to intervention may have distinct health consequences; some consequences may even be unexpectedly adverse. The task of social epidemiology is to sort out the wheat from the chaff. Arguably one of the more marked developments in the past decade of research on social determinants of health has been a flourishing of innovative strategies to evaluate causal hypotheses. The focus on distinguishing causal from noncausal associations goes hand-in-hand with the emphasis on "translation" from research to interventions to improve health.

In the first edition of this text, extensive evidence documenting social inequalities in health was presented (9). This evidence was roughly as strong as the evidence from observational epidemiology for *any* risk factor: associations were documented repeatedly, across diverse contexts and populations; and associations typically showed a dose-response relationship, such that higher SES was associated with better health across the spectrum of SES. A novice who is initiated into the discipline is likely to be overwhelmed by the sheer volume of evidence linking SES to health. The literature appears to be highly robust and consistent over history, stages of the lifecourse, and societal context and across measures of morbidity and mortality. Consistency, however, might be misleading: We may repeatedly find that people who are in the hospital have a higher risk of dying than those who are not in the hospital, but we would be mistaken to conclude from this consistency that hospitals kill people.[2] Likewise, the fact that SES is correlated with health in many countries and time periods tells us little or nothing about the causal nature of this association. Something that explains everything under the sun tends to arouse suspicion, and suggests that we may be ignoring important details. For instance, higher income is associated with greater likelihood of using a seatbelt while driving or going to bed at regular hours (10). Yet neither of these activities costs money, suggesting it is not the money per se that is promoting these healthy habits. Cigarette smoking is strongly correlated with educational attainment. But when we take a sample of adults with different levels of educational attainment and "look back" to when they were all the

2 Actually some hospitals do kill people—see, for example, the hospital scorecard developed by the Joint Commission—but that's another chapter in a different book.

same age (seventeen) and were still in school, with the same level of education, we find that the "educational gradient" in smoking was *already* apparent at that age (11, 12). This suggests that forcing people to stay in school and graduate from college may not prevent them from smoking. Maybe causality runs in the opposite direction, that is, smoking increases the chances of getting suspended, and eventually dropping out of school; or perhaps smoking and schooling share a common cause, for example, conflict with parents may be associated with both smoking uptake and poor school performance.

THEORETICAL DEBATES IN SOCIAL INEQUALITIES RESEARCH

Researchers have long engaged in heated debates regarding which dimensions of social disadvantage are most relevant for health, and the fundamental explanations for the existence of social inequalities. However interesting these debates may be, they can seem somewhat disconnected from public policy. To guide public health, we need not choose between Marx and Weber, but merely identify exposures that, if changed, will improve the health of the population. However, theories of social inequalities are important to the extent that they explain why certain intervention strategies are likely or unlikely to have substantial health impact in particular populations. Even interventions demonstrated to be effective in one population might not succeed in new communities or with seemingly innocuous modifications to the intervention strategy. Social theory—about what types of disadvantage matter for health and what mechanisms link social disadvantage to health—helps us understand how to generalize to new populations and how to predict beyond the scope of currently available intervention data. One of the most widely applied theoretical frameworks for understanding socioeconomic inequalities in health is Bruce Link and Jo Phelan's "fundamental cause theory."

SOCIOECONOMIC STATUS AS FUNDAMENTAL CAUSE OF HEALTH INEQUALITIES

According to Link and Phelan (8), high-SES individuals possess a variety of resources "such as money, knowledge, prestige, power, and beneficial social connections that protect health no matter what mechanisms are relevant at any given time." The theory has proved quite useful for understanding the persistence of health inequalities over time. Even as the major causes of morbidity and mortality changed over the course of the twentieth century, social inequalities persisted. As new diseases emerge, new inequalities emerge. As new preventive measures or cures are developed, inequalities in use or access to those innovations are established (13–16). Link and Phelan proposed that this was because low SES placed people at "risk of risk." Regardless of the specifics of the most salient disease risk factor in any given time or place, low-SES people are more likely to be exposed because high-SES people use their money, knowledge, prestige, power, and networks to avoid such exposures. Link and Phelan powerfully argued for the importance of evaluating not

only individual-level risk factors for disease, but also why some populations were more likely to be exposed to such risk factors than other populations. Further, they argued that focusing exclusively on the mechanisms that might link SES and health, rather than on SES itself, had several unfortunate consequences. First, disregarding the influence of social context on the distribution of risk factors may lead to interventions that are ineffective because they seek to change behaviors that are largely a consequence of factors outside the scope of the intervention. In other words, individuals are constrained to behave in unhealthy ways by their social circumstances, which render other behaviors extremely costly or impossible, and no amount of health information or good intentions will induce a long-term change. Further, Link and Phelan argue that focusing on individual-level risk factors proximal to disease may lead to blaming individuals for phenomena outside their control. For example, they note: "Morbidity and mortality due to tobacco is attributed to an individually-based bad habit rather than to a heavily advertised, government-subsidized, highly profitable killer industry" (8: p. 90).

A limitation of fundamental cause theory is that it does not imply anything specific about the linkages between different resources (knowledge, money, prestige, beneficial connections) and specific health outcomes. In other words, the theory does not help us to predict the marginal impact of changing one specific component of SES (e.g., income) on health. What it tells us is that the high-SES groups will tend to do better no matter what. In an ideal world, we should expand the access of low-SES groups across a broad swath of resources—schooling, income, safer jobs—but that insight does not help the policymaker to prioritize investments, nor does it offer a guide to predicting when specific policies are likely to generate unintended consequences. For example, consider the two following observations: *(1)* there is an inverse relation between income and smoking, that is, the higher a person's income, the less likely the person is to smoke; yet, *(2)* a short-term increase in income (for example, from winning a lottery) tends to lead to more cigarette consumption. How can we reconcile these apparently contradictory findings? The economic theory of the income elasticity of demand for cigarettes is consistent with the latter finding, namely, more income results in more consumption, all other things being equal.

To account for this paradox, within the fundamental cause theory, one might suggest that even if someone's income is increased in the short term, they may not use the money to improve their health because they are lacking the knowledge or other resources (e.g., social connections that discourage bad behavior) available to high-SES individuals. Socioeconomic inequality is sometimes likened to a flowing river; even if you could dam one or two branches of it, the water inexorably finds its way around it to flow downstream. As a description of the enduring nature of socioeconomic inequality, the fundamental cause theory provides a cautionary tale about the limits of social policy to reverse disparities. In other words, even if you could increase someone's income by a little bit, low-SES groups may still lack other resources such as knowledge, prestige, power, and beneficial social connections that are necessary to translate the increased income into better health. Conversely, high-SES groups manage to avoid health threats because they are able to use their access to resources in a flexible manner. For example, when they experience an increase in wages, they are less likely to use the windfall to purchase cigarettes (because they possess the knowledge that cigarettes are unhealthy), and more likely to save the money for retirement.

Despite the power of fundamental cause theory, it leaves several empirical observations unexplained, and is not sufficiently specific to guide development of interventions within the context of ongoing social inequalities. One limitation may emerge because for many, if not most, people,

pursuit of optimal health is not the preeminent goal to which they devote all of their resources. Other outcomes, such as psychological well-being, social integration, physical comfort or pleasure, or well-being of loved ones, may trump personal health goals (14). Thus, socioeconomic resources may often be put to work achieving other goals, even at a cost to health. This is especially relevant when the health outcome under consideration occurs late in life but the relevant risks are incurred early in life. One hypothesis about why education affects so many health outcomes is that it increases the individual's "time-horizon," and thus the relative importance of long-term health goals over short-term ambitions (see Chapter 13 on behavioral economics for further discussion). A related challenge is that "health" is not a unitary construct that would be promoted by a single, specific set of actions. Although many risk factors influence multiple health outcomes (e.g., cigarette smoking), some behaviors that are healthful in some contexts are unhealthful in others (e.g., eating calorie-dense foods). Finally, very little is known about how to prevent or treat many diseases. In such situations, the resources of high-SES individuals may have limited benefit (16).

A DYNAMIC APPROACH TO CONCEPTUALIZING SOCIOECONOMIC STATUS INEQUALITIES IN HEALTH

Underlying the fundamental cause theory and most classical social epidemiology models is the notion of a static, unchangeable SES dimension that is granted to us almost at birth and that persists throughout the lifecourse. In contrast, recent life-cycle models conceptualize SES as dynamic (17, 18); SES is not static, but changes as individuals move through multiple stages in the life cycle. For example, an individual may experience an "income shock" that profoundly changes consumption either temporarily or permanently. Similarly, governments may expand educational opportunities for entire cohorts, who may increase their educational attainment to levels never experienced by previous cohorts; or legislation affecting pension programs may increase or decrease the income of older people, leading to changes in consumption at older ages. Even if there were a static, unchangeable SES indelibly stamped on us from birth, it would be implausible to suggest that shocks to income and education have no impact on the consumption of goods or behavioral choices people make, many of which may be relevant to health. Within this dynamic model, individuals may use their socioeconomic resources to improve their health, consistent with "fundamental cause" theory, but they may also use their resources for other purposes, some of which may be health-damaging. Similarly, people may take advantage of their good health to improve their social resources, for example by accepting hazardous but high-paying employment.

This dynamic approach also has major implications for how we conceptualize the causal nature of SES inequalities in health, because it implies that health and SES are reciprocally, dynamically affected by each other (17, 18). A major adverse "health shock" during the early years of life, for example, a chronic disease diagnosis or major injury, may prevent a child from achieving the educational level he or she would have otherwise attained. Similarly, a new diagnosis of a chronic disease may hamper the ability of a middle-aged worker to accumulate income and lead to depletion of financial assets to fund consumption or healthcare. The death of a spouse may lead to both poorer health and lower income, generating an association between SES and health. In order to disentangle the multiple causal and noncausal mechanisms generating SES inequalities

in health, we therefore need a less restrictive model that is dynamic and incorporates the possibility of mutual dynamic influences between health and SES over the life cycle. Conceptualizing SES inequalities in health in this dynamic framework is consistent with our key interest, as public health practitioners, in identifying whether interventions that change SES may potentially influence population health.

SOCIAL CLASS, SOCIAL POSITION, OR SOCIOECONOMIC STATUS?

Traditional social stratification theory seeks to explain the existence and persistence of social classes. For example, Marxist grand theory posits that capitalism generates two types of classes (i.e., the bourgeoisie and the proletariat) defined according to their relation to the means of production. Alternatively, in the Weberian tradition, classes are defined by multiple dimensions besides their relationship to the means of production, such as income, status (honor, prestige), and power.[3] In contrast to these categorical approaches to defining social classes, social epidemiology (as well as other related fields such as psychology and economics) tends to adopt a more flexible "gradational" approach to understanding social inequalities (19). Rather than focusing on sharp boundaries between social classes, the most relevant determinants of health correspond to continuous indicators such as income, wealth, human capital (education), status, or prestige. Health inequalities are expected to emerge across the continuum of such indicators.

The social epidemiological orientation is motivated by a desire to understand *what can be done* practically in the realm of SES and health. Social inequalities in health differ substantially across time and place, as do racial inequalities (20–25). This variability provides compelling evidence that inequalities in health are not inevitable. A range of social and policy actions can likely be used to remediate inequalities and improve the health of the most socially disadvantaged. Thus, if some aspect of SES is truly causal for health, what actions should we advise policymakers to take in order to tackle health inequalities? To identify the most effective strategies to address health inequalities, it is necessary to "de-compose" SES as a categorical concept, and to drill down separately to the bedrock of relationships between its constituent parts and health outcomes. To wit, social epidemiology seeks to generate actionable information for policy translation.[4] For example, if we agree that some part of the relationship between schooling and health is causal, then we need to understand where society should invest—by subsidizing preschool programs, or encouraging

3 For an exhaustive (and occasionally exhausting) survey of contemporary debates about social class, see David Grusky, ed. *Social Stratification in Sociological Perspective.* Boulder, CO: Westview Press; 1994.

4 The study of SES in social epidemiology does not seek to make any predictions about the "historical laws of motion" of capitalist society *qua* Marxist grand theory. SES is of primary interest to the social epidemiologist because differences in access to its constitutive resources—namely, income, wealth, human capital, social status, occupational prestige, and authority—also generate *health* stratification. The social epidemiological approach thereby discards questions about class generation and reproduction, class consciousness, or class struggle and action; epidemiologists are usually content to leave those kinds of questions to the sociologists. For the foregoing reasons, the work of social epidemiologists is sometimes attacked as being crassly empirical and "atheoretical." Our chapter might be criticized for focusing too narrowly on causal relations between income, education, and health. In doing so, we have ceded the broader (and deeper) questions about the relations between social class and health, as well as the political economy, welfare regimes, democracy, and so on.

high school graduation, or expanding access to community colleges. If we agree that some part of the relationship between income and health is causal, then we need a better understanding of how different types of income transfers might improve or harm health. For example, cash transfer programs often condition payments on certain behaviors, such as taking children to school or participating in vaccination programs. These conditional programs may influence child health outcomes in fundamentally different ways than income transfer programs that have no strings attached, or require parents to engage in employment, leading to lower parental inputs on children. A key question for social policy is how to minimize unintended consequences of making cash transfers to the poor, so that they spend their money to maximize their welfare (for example, using food stamps to purchase nutritious food for their children, or conditioning cash transfers on behaviors with long-term health and social benefits).

Socioeconomic status and socioeconomic position (SEP) are often used interchangeably in the epidemiological literature—or at least, many authors who use the term "SEP" seldom take the trouble to spell out their theoretical orientation. Strictly speaking, they should not be used interchangeably. Socioeconomic position is a *relational* concept, encoding how groups stand in relation to each other—for example, in the context of a workplace, there are those who occupy positions of supervisory authority over subordinates versus those who take orders from above.[5] Contrasting with this theoretical orientation, socioeconomic "status" is used as a term to refer to *differences* between individuals and groups in the possession of resources—for example, differences in schooling, income, or occupational prestige—without necessarily attributing any causal connection between the status of one individual vis-à-vis another. For example, suppose we observe that the owner of a textile mill makes many times more income than his workers. From a relational perspective—for example, a Marxist standpoint—these inequities arise as a consequence of class relations, that is, the owner is wealthy because his class position enables him to commandeer the labor of the workers and provide little in exchange. We are not adopting a framework of class analysis in this chapter, so we shall stick to the terminology of SES. The reason for this choice is that, as explained earlier, such frameworks do not enable us to derive direct implications on how to intervene in specific components of SES in order to improve health. The rest of this chapter focuses on two dimensions of SES, namely, schooling and income. The third—occupation—is to a large extent addressed in Chapter 5, which examines how factors linked to one's occupation, such as the level of flexibility and control over work, might influence health. For a detailed guide to the measurement of occupational status, the reader is also referred to Chapter 2 of the first edition of *Social Epidemiology* by John Lynch and George Kaplan (28).[6]

5 The concept of class position aligns with what Simmel called "empty spaces" (*leere Raum*), that is, they create inequality independently of the characteristics of individuals who end up filling the positions. It follows that only by changing the class structure can the structure of inequality be changed (26, 27).

6 In North America, there are fewer studies of social stratification *in health* based on occupational classification compared with studies based on income and educational differences. This contrasts with the UK, where the Registrar General's classification of occupations ("social classes") on the Census was the basis for government monitoring of health inequality from 1911 until it was discontinued in 2001 and replaced by a new measure—the National Statistics Socio-economic Classification (NS-SEC), based on employment relations and conditions of occupations (29). The UK is also home to the Whitehall civil servants studies, which were seminal in the "discovery" of the socioeconomic gradient in health. Since everyone in the British civil service is assigned to an occupational classification (e.g., administrators, professionals, clerks, janitors, tea-ladies, and so on), it seemed natural that the study would focus on occupational grade as the primary measure of SES.

There has been substantial debate about the most relevant measures of SES for specific health outcomes, for example, education versus income versus occupation. There are two fundamental problems with an approach that focuses on identifying the "best" single measure of SES: This pursuit is guided by the misconception that different measures of SES are just manifestations of an underlying SES dimension, and as such our only task is to figure out which of the three measures more closely reflects that unobserved, latent SES dimension. As we shall see, although different SES domains are interrelated (education affects income, income affects wealth), each of them constitutes a different set of health-relevant resources, and their effects on health may be fundamentally different. Second, an approach that drills down to specific SES measures is necessary for defining interventions and informing our theoretical understanding of mechanisms through which SES affects health. For example, for cognitive outcomes, we might anticipate that education is a more salient SES measure than wealth. On the other hand, direct measures of financial resources may be more relevant for health conditions closely related to access to medical care, such as chronic disease management. We focus on education and income because both are easily amenable to interventions and there is extensive research on both.

SOCIOECONOMIC STATUS AND HEALTH ACROSS THE LIFECOURSE

In order to identify effective approaches to address social inequalities in health, it is critical to incorporate the dimension of time and, in particular, the differential influence of stages of the lifecourse. This entails considering when an exposure causes disease—that is, when it becomes "physiologically embedded" such that removing the exposure does not eliminate the harm the exposure has already caused. Some developmental stages may be more "sensitive" to context, so exposures during that period have larger effects on health than exposures that occur either before or after that window. Such sensitive periods may be defined by physiologic (rapid cell growth), psychological (attention to peer behaviors), or social (entry into the labor force) events. Time is also relevant for understanding when it is most feasible to change an exposure. For example, although much of the adverse health effects of smoking accumulate over decades, smoking initiation typically occurs within a relatively narrow age band in adolescence.[7] Thus, although changes in smoking behaviors might be healthful throughout life, interventions to prevent smoking are likely to be most effective in adolescence.

Conventional lifecourse research considers three accounts of how the timing of social exposures, such as low SES, influences health (summarized in Figure 2.1) (31). *Critical*, or *sensitive period* models suggest that exposures during a specific developmental period have reverberating consequences years later that cannot ever be fully ameliorated. *Accumulation of risk* models imply that each additional episode of low SES adds to an ever-growing health disadvantage. *Trajectory*, or *chain-of-risk* models posit that low SES is primarily unhealthy because it begets future low SES, and only in later life does low SES become biologically embedded as ill health. In addition to these classic three models, some exposures follow an "immediate risk" model, such that the link between social exposure and adverse health outcome is effectively immediate. Finally, there is

7 Over 90% of lifelong smokers report that they started smoking before age 18 (30).

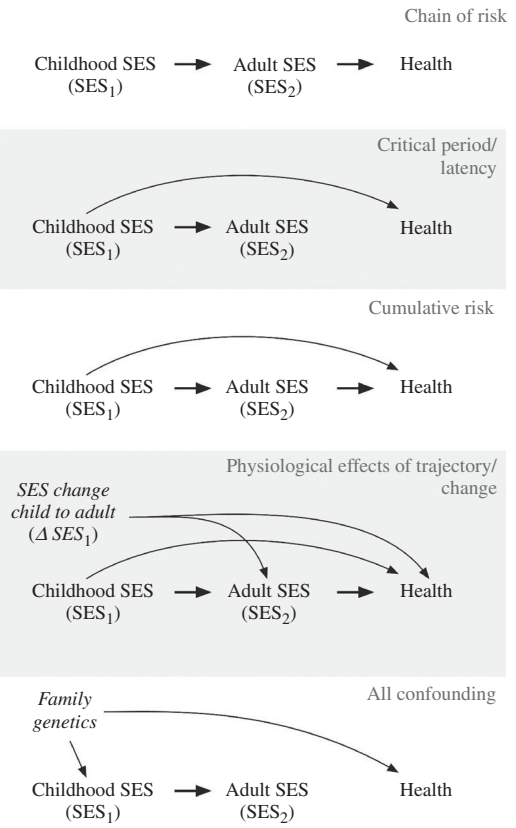


FIGURE 2.1: Alternative lifecourse models of etiologic periods linking SES and health.

some evidence that instability in SES per se is important for health, over and above any effect of disadvantaged SES at one point in time.

Distinguishing between these models is valuable because they guide interventions. On the one hand, if there is a critical period in early life, interventions targeting adults are a waste of time and energy. On the other hand, under an accumulation of risk model, early intervention might be best, but later intervention could also be valuable. These models are empirically distinguishable (32), although efforts to formally test models against one another to date have been hindered by measurement error and time-varying confounding problems (33–35). Note that there is no reason the same model should apply to all SES-health outcome combinations. In Chapter 14, on physiologic pathways linking social adversity and health, we note various mechanisms that may correspond with either sensitive periods or accumulation of risk models.

The SES-instability model can explain some empirical puzzles in the SES and health literature. For example, virtually all studies agree that downward social mobility predicts bad health (admittedly, some of this is due to "reverse causation," or downward drift, which we discuss further below). More surprisingly, a few studies also suggest that *upward* social mobility can be detrimental to health. How come? The answer to that question may depend on the *time frame* over which upward mobility is defined and measured. Over short time periods (e.g., a few months), an

increase in income may produce more unhealthy behaviors (see previous example of wage increases and increased cigarette consumption). Even when social mobility is defined over a single person's lifecourse (e.g., low parental income during childhood versus high income attained during adulthood), there may be an adverse effect of instability. How can this be explained? In Chapter 14, we describe a set of biological embedding mechanisms that involve a predictive adaptive response early in development, based on the *anticipated* adult environment. Such adaptations may be either physiologic (e.g., altered glucocorticoid receptor expression programmed by early life maternal interactions) or behavioral (e.g., more frequent externalizing behaviors in response to living in a highly regimented institutionalized environment), but they are adaptive within a specific context. Outside that context, however, or over the long term, they may be unhealthy. For example, one hypothesis is that individuals who were nutritionally deprived in utero follow a developmental track that improves survival in calorie-poor environments, but increases risk of cardiometabolic problems in calorie-rich environments. Such a mismatch between fetal environment and postnatal circumstances can explain why changes in SES might be harmful. This is related to the more psychologically based "status inconsistency theory" (à la Robert Merton) that people from low-SES backgrounds don't "fit" in high-SES society (the "Eliza Doolittle effect"), and that this may lead to stress and bad health outcomes. There are some documented examples of this in recent research; for example, in a longitudinal study of 102 adolescents, Marin et al. found that a socioeconomic trajectory starting with low early-life SES that increased through childhood was associated with the highest blood pressure levels in adolescence (36). In another follow-up of a representative sample of 489 African American youth living in the rural South, those adolescents who exhibited high levels of competence at ages 11–13 (as reported by their teachers on ratings of diligence, patience, and social skills) were more likely to be enrolled in college at age 19. However, the same upwardly mobile children were also more likely to have higher indicators of "allostatic load" (see Chapter 14), as assessed by higher body mass index, higher blood pressure, and stress hormone profiles (cortisol and catecholamines) (37). The authors interpreted their findings as a corroboration of Sherman James's John Henryism hypothesis (38), that is, active striving to achieve upward social mobility can lead to deleterious health consequences for disadvantaged groups, in this instance, rural black Americans. Of course, there is also plenty of evidence that upward social mobility is by and large good for health; nonetheless, a comprehensive theory of SES and health across the lifecourse must be able to accommodate (and explain) contrapositive findings as well.

A COUNTERFACTUAL FRAMEWORK FOR A PRACTICABLE SOCIAL EPIDEMIOLOGY

As we have already intimated, it is important that social epidemiologists focus on *causal inference* in order to guide health interventions and policy translation (39). Counterfactual accounts of causation define the causal effect of one treatment or exposure versus another as the contrast between the outcome an individual exposed to the treatment actually experienced, and the outcome the same individual *would have had*, if (counter to fact) she or he had instead experienced the other treatment. For example, in 2006, life expectancy at age 25 for men with a high school diploma or General Educational Development (GED) was 4.3 years longer than for men who had

not completed high school; for women, this gap was 5.3 years (40). To describe this association as causal is to say that if we could somehow go back in time and induce the previous dropouts to complete high school, the men would on average survive 4.3 extra years. This definition of causal effects highlights four points of special relevance to social epidemiology:

1. Many statistical associations between exposures and health outcomes are probably not due to the causal effect of the exposure on the outcome, but rather due to reverse causation or confounding.
2. The measurement of a social exposure might correspond with a great diversity of actual experiences, each with distinct health consequences. The health effects of graduating from one of the Saint Grottlesex prep schools likely differ from the health effects of graduating from an underfunded, segregated public school.
3. Different mechanisms of achieving the same "exposure" may have different health consequences. High school attendance induced by financial subsidies and appropriate social supports might have different consequences than high school attendance mandates enforced by truancy officers.
4. An identical exposure, induced by an identical mechanism, may nonetheless have different effects on different people. Some people do not benefit much from a Saint Grottlesex education, and other people capitalize brilliantly on the public school diploma.

Not all relationships between SES and health outcomes are likely to be causal (41). One potential misgiving about the application of a counterfactual framework in social epidemiology is that the causal systems determining health are so complex—thick with feedback loops and nonlinearities and interactions—that it is impossible to isolate causal exposures.[8] In this account, it is impossible to say that it is high school completion *specifically* that makes the difference for life expectancy. Although there is a certain appeal to throwing up one's hands and saying "this is such a complex system that those simplistic causal questions are irrelevant," doing so also abrogates any relevance to improving public health. If our goal is to advise individuals, clinicians, or policymakers about what actions to take to improve health, then we must be willing to make causal claims about specific interventions.

8 A famous, albeit slightly morbid, puzzle describes a man lost in a desert with two companions. The two companions each decide independently to murder the man (the reason the man is so generally despised varies across incarnations of the puzzle). One of the companions sneaks cyanide into the man's water canteen, so the man will be poisoned to death when he drinks the water. The other companion, unaware of the poison, punches a hole in the canteen so the water leaks out during the night. Both companions sneak away and leave the man, who shortly thereafter dies of thirst wandering in the desert. The puzzle is: which, if either, companion murdered the man? The man never drank the poisoned water, so it can hardly be argued that the poisoner killed him. The water that leaked out of the canteen was poisoned and would have killed the man had he drunk it, so the companion who punched the hole in the canteen may also plead innocence—if anything, he prolonged the man's life by preventing him from drinking cyanide-laced water. This relates to the debate about the challenges of isolating social causes of disease because some groups may be so disadvantaged that alleviating one adversity will not be enough to enable them to achieve good health. The bad health outcomes are "overdetermined" by multiple sufficient causes of disease. In a counterfactual analysis, removing only one of the sufficient causes would suggest that it was not a "cause" at all, because the disease outcomes did not change. From the perspective of how to intervene to improve health, this is the right answer: changing this one factor is not enough to improve health.

As a consolation to the "we can't simplify" camp, it *is* often useful to identify a "bundle" of interventions that, when delivered together, are health-promoting, even if we do not understand which specific component of the intervention is helpful. For example, conditional cash transfer programs, described below, deliver a bundle of treatments. The finding that this bundle improves an array of health outcomes is extremely useful, even if we are not certain whether the most important aspect of the bundle is the extra money, or the conditions required to receive the extra money, or some other aspect of the program. Isolating the most beneficial ingredients of a bundled package is not necessarily essential to guide policy, and often simply establishing the effects of the bundle is very helpful. However, identifying the most beneficial ingredients is useful because such knowledge gives policymakers better tools to design new programs and to predict generalizability of findings (42).

It helps to be able to understand what aspects of SES we ought to intervene on (and at what stage of the lifecourse) in order to address health inequalities, as well as to be able to anticipate when (and for whom) interventions are likely to be useless or even harmful. Randomized controlled trials (RCTs) are the gold standard to establish causality because they can rule out the possibility of reverse causation or confounding. Implicit in the implementation of an RCT is a clear specification of who is to be enrolled, what the exposure is, and how the exposure or treatment is to be delivered. No doubt RCTs have several limitations, however. Results from an RCT may not be generalizable to people who did not participate in the trial. In addition, a program evaluated in an RCT may sometimes show a different effect when scaled up, due to equilibrium effects or spillovers. This is of special relevance to public health problems, where outcomes depend on collective action, and arise from a complex multiagent interaction not easily reproducible in experiments. An implication is that behavioral responses induced in RCTs may not always correspond to real people's nonrandom responses following the introduction of a program or policy (43). Despite these limitations, randomization remains a powerful tool in forming an evidence base for a practicable social epidemiology. Whenever it is impractical to conduct trials, or when general equilibrium effects may be a concern, natural experiments offer a promising alternative to identify causal relationships using observational data. When we adopt an explicit counterfactual framework, our studies are more likely to yield useful insights for a practicable social epidemiology.

THREATS TO CAUSAL INFERENCE: REVERSE CAUSATION

Researchers have long recognized reverse causation as a threat to causal inference in observational studies of SES and health. The seminal Black Report on Health Inequalities (44)—which represents one of the first systematic attempts in the modern era to understand the causes of health inequalities—acknowledged that at least some portion of the association between SES and health likely reflects "downward social drift." For example, the strong correlation between SES and mental health cannot be fully explained by the fact that socioeconomic disadvantage and adversity cause anxiety, depression, and psychological distress. Rather, the onset of mental illness is a potent trigger for loss of employment and earnings, as well as out-of-pocket medical expenses. Ignoring reverse causality in this instance will result in an overestimation of the relation between SES and mental health.

In the conventional wisdom on SES and health, it is often taught that education (as an indicator of SES) is less susceptible to problems of reverse causation, compared with indicators such as income or occupation (41). The reason is that most people have (fortunately) completed their schooling by the time they begin to fall prey to chronic diseases. And once you get sick, although you can lose your job and income, you cannot lose your education—that is, the men in white coats cannot take away your diplomas after you are diagnosed with an illness.[9] But is this account strictly accurate? As a matter of fact, the more closely we look at this question, the more evidence we find that reverse causation is real, even in the case of schooling and health. Thus, Anne Case and colleagues (45) turned to evidence from the 1958 British Birth Cohort (the National Child Development Study), and found that chronic health conditions during childhood do appear to have an adverse impact on educational attainment. Even after taking into account household and parental characteristics, each chronic condition reported at age 7 led to an average 0.3 fewer subjects passed on General Certificate of Education O-level examinations at age sixteen among children in this cohort. In short, chronic conditions during childhood—for example, juvenile-onset diabetes, severe asthma, ADHD, or mental health problems—result in children missing school. Similar evidence in the United States suggests that childhood obesity influences school completion rates (46).

Turning to income and health, there is again evidence in support of reverse causation for certain outcomes. For instance, the presence of a gradient relationship between income and overweight or obesity has been repeatedly observed. What is often unremarked (or unnoticed) is that the inverse relation between household income and overweight/obesity is confined to one gender, namely, women (at least in US data). Among men, the relation between income and overweight is flat—or even tends to run in the *opposite* direction. For example, NHANES surveys indicate that in 2005–2008, (47) the prevalence of obesity among men with household incomes at or above 350% of the poverty level was 33%; whereas among men living in households with income below 130% of the poverty level, obesity prevalence was 29%. There is something decidedly "fishy" about the idea that lack of income causes women to be overweight, but men to become leaner. If lack of income is supposed to lead to poor nutritional habits, why is the income-obesity association only found in women?

The likely answer is that the relation between income and overweight/obesity is partially driven by reverse causation, that is, it is not that lack of income causes people to gain weight; rather, it is overweight/obesity that leads to loss of income. Experimental and observational evidence suggests weight-related discrimination is much worse for women than men (48, 49). If reverse causation is the key phenomenon accounting for the association between low income and overweight status, it would not be surprising that it emerges only among women. Based on a longitudinal analysis of the Panel Study of Income Dynamics, Conley and Glauber (48) tested this reverse causation hypothesis among adults aged 25 years and older. The result of this analysis showed that each 1% increase in women's BMI at baseline in 1986 was associated with 0.6% lower family income, 1.1% lower spousal earnings, as well as 0.3% lower probability of being married during fifteen years of follow-up. In other words, overweight and obese women face more difficulties in competing in the labor market and the marriage market. Moreover, this phenomenon *differs*

9 As in Philip Larkin's poem "Days": "What are days for?. . . Ah, solving that question/Brings the priest and the doctor/In their long coats/Running over the fields."

by gender and race, as there appeared to be no association between baseline BMI and subsequent economic or marriage outcomes among men or black women in the same study. That is, the association between income and overweight/obesity is likely to be explained at least to some extent by societal "fat bias" operating against white women (but not white men). A corollary implication of this study is that raising the incomes of poor women may do little to help them lose weight; and indeed, an analysis of income transfers via the Earned Income Tax Credit (EITC) suggested that raising incomes led to *weight gain* among female recipients (50). This is an illustration of when adopting the counterfactual framework can guide policy; that is, just because there is a correlation between low income and overweight/obesity, it does not necessarily mean that raising the incomes of the poor will help to produce weight loss.

THREATS TO CAUSAL INFERENCE: CONFOUNDING

Another major threat to causal inference is confounding, that is, the notion that a correlation between income or schooling and health is driven by the common prior influence of an omitted (unobserved) third variable. Once again, a close examination of the data reveals that not all parts of the association between SES and health are driven by forward causation. For example, there is a strong relation between educational attainment and smoking status. Yet as far back as 1982, Farrell and Fuchs (11) noticed that the "educational gradient" in smoking was already apparent among individuals before they completed schooling. In other words, they showed educational inequalities in smoking in a sample aged 24, and then demonstrated that those differences were *already present* when the same individuals were only aged 17 (and when everyone had the same years of education). Readers might object that there may have been differences in the quality of education up to age 17. But that would seem an unlikely explanation, given that additional years of education beyond age 17 did not alter the gradient; in other words, one would have to argue that quality of education mattered for smoking only up to grade 12 but not beyond that. A more parsimonious explanation is that schooling is not causally related to smoking initiation;[10] rather, the association is likely to be explained by unobserved third variables that underlie both educational achievement and smoking initiation. An example of such a third variable might be differences in delay discounting between individuals. As we shall explain in the chapter on behavioral economics, there are individual variations in the ability to delay gratification and to invest in the future. More "patience" and self-regulation is likely to be a common prior cause of how long youths stay in school, as well as whether they succumb to tempting habits such as smoking.[11] The lack of a causal association between schooling and smoking initiation is further buttressed by studies such as the twin fixed effects analysis conducted by Fujiwara and Kawachi (51). In the MIDUS sample of monozygotic twins (who shared the same genetic background as well as early rearing environment), there was no relation between educational attainment and smoking rates, once the investigators took account of time-invariant confounding factors. Weak associations between schooling

10 Of course, the results of this study do not negate the possibility that achieved years of education might make a difference to smoking *cessation* rates later in life. This simply underscores the point that social epidemiologists need to be very specific about the health outcome, that is, the predictors of smoking initiation may differ from the predictors of quitting.

11 We hasten to add that education can strengthen self-regulation and lower people's discount rates. At the same time, there is also likely to be an inherited component to delay discounting as well.

and smoking uptake have also been observed in sibling fixed effect models that control for unmeasured familial vulnerability to smoking (12).

APPROACHES TO IDENTIFYING CAUSAL EFFECTS IN SOCIAL EPIDEMIOLOGY

Research on social determinants of health is conducted in several disciplines, including epidemiology, economics, sociology, and psychology. Drawing on the research modalities of these other disciplines has been invaluable to helping evaluate causal questions. Table 14.1 summarizes research designs commonly used in our field, and some strengths and weaknesses in each. With respect to the effects of SES, there are a handful of truly randomized experiments available to us. In some cases the randomization was "clustered," but a clustered randomization design is no less rigorous. Because equipoise is not plausible when the treatment in question would lift a family out of abject poverty, some of the most important randomized trials follow a staggered or wait list design: everyone will eventually get the treatment but resource limitations dictate that some people are randomized to receive the treatment earlier while others will receive it later. Again, the design is rigorous but it can only identify the effects of the short-term difference in treatment, not the cumulative effect of long-term exposures.

Randomized trials, however, are difficult to conduct and therefore quite rare. Substantial emphasis is thus placed on findings from quasi-experiments or natural experiments, what could be called "pseudo-randomized" trials. There is a close conceptual relationship between several research designs, variously called "instrumental variables" (IV), difference-in-difference, and regression discontinuity designs applied to understand the effects of an exposure on an outcome. These approaches are not yet standard in epidemiology, but they are the workhorses in many areas of economics research (52). All of these, IV, difference-in-difference, and regression discontinuity designs, rely on the assumption that an exposure of interest differs or changes for reasons that have nothing to do with the outcome of interest. Each of these approaches should be evaluated with the same skepticism about that assumption.

We briefly describe IV models here, because they have been used extensively in research on SES. For example, understanding of the effects of education has been influenced by several studies that use changes in compulsory schooling laws (e.g., mandating the minimum school dropout age) set by state or national governments as natural experiments to estimate the effects of additional years of education on adult health. Most children do not drop out of school as soon as it is legal to do so; some drop out earlier regardless of the law, and many continue on for years after the law allows them to quit. This means that the average effect of the laws on all children was very small. Angrist estimated that for the 1944 US birth cohort, dropout at age 16 was 4 percentage points lower for states that mandated schooling until age 17 or 18 than in states that permitted dropout at age 16 (53).

To motivate discussion of IV analyses, we use a causal directed acyclic graph (54) (DAG; Figure 2.2). Imagine we wish to estimate the effect of an exposure or treatment X on a health outcome Y, but we fear there may be one or many unmeasured factors that influence both X and Y (represented by "U" in the DAG). Because these variables have not been measured, they are of necessity omitted from our regression model, inducing a bias that economists attribute to "omitted variables," and

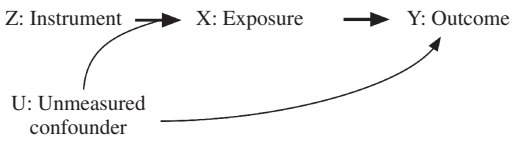


FIGURE 2.2: Causal structure assumed in instrumental variables analyses.

epidemiologists attribute to "confounding." Fortunately, we have identified a variable "Z" that influences our exposure of interest, but does not directly affect the outcome, and shares no common causes with Y. Z is considered an "instrumental variable" for the effect of X on Y (55). Referring to the figure, the key assumptions are *(1)* the *presence* of an arrow from Z to X, *(2)* the *absence* of a direct arrow from Z to Y, and *(3)* the *absence* of any variable that has arrows into both Z and Y.

Epidemiologists may recognize that this diagram matches exactly the structure of a randomized controlled trial, in which Z represents random assignment, X represents treatment received, and Y is the outcome of interest. It is taught in Epi 101 that all RCTs are evaluated with an "intent-to-treat" design: The association between random assignment (Z) and the outcome is used to test the null hypothesis that X has no effect on Y. We do not use the association between treatment received (X) and the outcome to assess this null hypothesis because the possible presence of factors U that influence adherence to randomly assigned treatment and the outcome would introduce an association between X and Y even if there were no causal effect of X on Y.

The intent-to-treat estimate tests the null hypothesis of *no* effect of X on Y, but if we reject the null and conclude that our treatment does influence the outcome, we still do not have an estimate of the magnitude of the effect. Many people randomly assigned to the treatment do not, in fact, adhere to their random assignment, and in the case of quasi-experiments, the association between "pseudo-randomized" treatment assignment and actual treatment is often very small. To the extent that Z does not perfectly determine X, the association between Z and Y will be a diluted estimate of the effect of X on Y. To estimate the effect of X on Y, IV analyses account for the attenuation due to weak effects of Z on X. The simplest IV estimator, applicable when Z, X, and Y are all binary, is simply the ratio of the association between Z and Y (the intent-to-treat effect estimate) divided by the association between Z and X (the "adherence" of participants to their randomized assignment). Because Z is randomized, both the numerator and denominator of this ratio can be estimated without bias. Intuitively, this is similar to a correction for nonadherence that one might apply in analyses of a randomized controlled trial.

There are some special caveats in most applications of IV analyses to data from quasi-experiments. Because the influence of the instrument on exposure is often very small, the estimates are typically very imprecise—the precision of IV effect estimates declines sharply as the strength of the association between the instrument and the exposure declines. Further, small violations of the assumptions, for example very small direct effects from the instrument to the outcome, can introduce large biases in the IV effect estimate. Finally, if the exposure does not have the same effect on everyone in the population—for example, if a year of education has a huge benefit for some people but a trivial effect on other people—then additional assumptions are necessary to interpret the IV estimate as the causal effect of X on Y for any particular subgroup of individuals. The most common interpretation is that the IV effect estimate refers to the effect of X on Y among

those individuals whose exposure X was changed by the instrument Z (56, 57). Thus, if we use compulsory schooling laws to identify the effect of an additional year of education on life expectancy, we are describing in particular the effect among those individuals who remained in school precisely because of the legal requirement to do so. Although this is often considered a limitation of IV estimates because there is no way to identify exactly who those people are, that population most influenced by the policy change may in fact be of special interest to policymakers.

EDUCATION AND HEALTH

Since Kitigawa and Hauser's landmark study on the association between education and mortality (58), there have been countless studies describing a relation between schooling and health outcomes. For example, the large National Longitudinal Mortality Study (NLMS) with follow-up from 1979 to 1985 suggested a 7–8% reduction in mortality per one-year increase in education among individuals ages 35–54 (59). Although relative effect estimates are smaller among older individuals, absolute effects are often much larger because of the higher death rate in older adults (60). Further, educational gradients in mortality are evident throughout the world, although the association between education and mortality tends to be smaller in some European countries, particularly for heart disease mortality (24). In the year 2000 in the US, education was associated with mortality rates for diseases of the heart, cancer, chronic obstructive pulmonary disease, stroke, and unintentional injuries among white men, white women, black men, and black women (61). The contribution of different causes of death to differences in mortality by educational level, however, differs considerably across different countries (21, 62).

Although the association between education and health is large and robust, this finding does not necessarily indicate that improvements in education would improve health and survival. To address this question, we turn to experimental, quasi-experimental, and especially rigorous observational studies of how changes in education influence health and mortality. Before discussing these, we begin with a seemingly more straightforward question of defining our exposure.

WHAT IS EDUCATION?

Nearly all research on education and health operationalizes education as either years of schooling, or degree credentials. The inadequacy of this conceptualization is increasingly evident. First, this ignores the tremendous variability in the *quality* of schooling between individuals. Manly has argued that literacy is a better measure of educational experiences than years of schooling, especially when comparing individuals likely to have experienced systematic differences in school quality, such as US blacks and whites. She finds that racial differences in old age cognitive outcomes, although not explained by differences in years of schooling, *are explained by measures of literacy* (63). These findings should be cautiously interpreted given the close link between typical old age cognitive assessments and literacy measures. Historically, there have been tremendous differences in both quantity and quality of education both between races (de jure racial segregation of schools prevailed in much of the US until 1954, and de facto segregation continues today),

across place of residence at both a local/neighborhood and regional level, and by parental SES. It is difficult to appropriately quantify these differences because it is difficult to measure "quality" comprehensively. For example, quality of schooling is likely influenced not only by formal resources such as expenditures or teacher training, but also by differences in peer groups. Nonetheless, even superficial measures of quality, such as the number of days in a school term an average student attended school, indicate the magnitude of differences in educational quality. Because of differences in term length (as well as even more extreme differences in the average number of days of class attended by a child in schools serving black children versus white children), the same number of "years" of schooling can translate into stark differences in the actual numbers of days that a child sat in a classroom. A typical school year in contemporary US schools is about 180 days, so we could consider 180 days of classroom time to be the equivalent of a school year. How large were the differences in "school year equivalents" by race and state in the early twentieth century? A white child born in 1925 who attended school in South Carolina for 10 calendar years averaged the equivalent of 2.6 more school-years of in-class time compared with a black child in the same state and birth year, but 1.0 year less than a New York child born in the same year (64).

A very small body of research has addressed the long-term health effects of school desegregation, reporting reduced adolescent fertility for black girls (65) and improved self-rated health in adulthood among blacks who attended schools in desegregated communities (66). Johnson also reports, based on sibling fixed effects models, substantial benefits to self-reported health of preschool attendance and district per pupil spending (66). Frisvold and Golberstein found that disparities in school quality measures for racially segregated schools partially accounted for racial disparities in disability, but not BMI, smoking, or self-rated health (67). We also know, for example, that early childhood school quality has enduring consequences for the child's adult earnings. The Tennessee Project STAR (Student Teacher Achievement Ratio) experiment focused not on increasing early education but on enhancing the quality of education typically provided to small children. In the STAR experiment, children were randomized to one of three groups: classrooms of 13–17 students, regular classes of 22–25 students, and regular classes of 22–25 students with a classroom aide. The intervention (discussed in more detail below) covered kindergarten to third grade. Because children were randomly assigned to a classroom, regardless of the class size, some children ended up in classrooms with excellent teachers and high-achieving classmates while other children ended up in classrooms with less stellar teachers and classmates. Without measuring these specific inputs, the researchers assessed "classroom quality" based on each student's classmates' end of year test scores (i.e., the classroom average, excluding the index student). Children randomly assigned to one of the high-performing classes had better college attendance rates and higher earnings; a 1-standard deviation improvement in "classroom quality" predicted $1,520 higher earnings at age 27 (68). Only a small fraction of the estimated effects of class quality was explained by measured features, such as teacher experience.

An additional shortcoming in the typical measures of education is that they do not incorporate nontraditional educational experiences, such as the General Educational Development (GED) credential, educational experiences directly related to work, or educational experiences pursued by many adults out of personal interest long after the completion of formal schooling. A handful of studies indicate that individuals with GEDs have health inferior to the health of high school graduates, and it is unclear if their health is significantly better than that of high school dropouts (69–73). It is possible that the GED is of greater benefit if it is completed at a younger

age, consistent with lifecourse theories, but this is not established. In 2004, approximately 44% of US adults who were not full-time students nevertheless participated in some type of educational activity, largely work-related activities and personal interest courses (74). Although these activities are common, average hours devoted to such adult education programs are generally a small fraction of the time commitment associated with formal schooling, so health effects may be modest. There is a large literature suggesting that leisure activities entailing cognitive and social engagement predict delayed onset of dementia, but because of the powerful influence of mental status on activities, causal inferences are particularly difficult to draw in this area (75).

Finally, typical measures of education pay little attention to the timing of educational experiences or preschool activities. Early childhood, because of rapid cognitive development, is a sensitive period in which educational exposures may have relatively larger impacts than educational experiences later in life (76). Thus, studying the long-run effects of early childhood education on health is a key area so far little explored in social epidemiology.

MECHANISMS LINKING EDUCATION TO HEALTH

To evaluate the credibility of studies linking education and health, it is helpful to understand the mechanisms through which education may influence health. First, education is the portal to occupational options (e.g., safer jobs) and higher earnings. Educational differences are sometimes dismissed as spurious on the grounds that it is *really* income that affects health. This reflects an incorrect understanding of causation. There is overwhelming evidence that, on average, attaining more education increases income (77). Therefore, to say that education does not affect health if income affects health is somewhat like arguing that vaccines don't prevent disease since it is really the immune response triggered by the vaccine that is protective. The implication is that policies that increase educational attainment have the potential to influence health, even if the effect is "mediated" by increases in income (which, by and large, is unlikely to be the only mechanism linking education to health).

There are several other mechanisms besides income through which education might influence health. Education may convey specific, factual information relevant for preventing disease or delaying disability and death after disease onset. Factual knowledge directly obtained from schooling seems unlikely to be a major mechanism for the health benefits of education, however. Most school curricula devote limited time to health education. The majority of lifelong smokers have already initiated their habit before their schooling is complete, and the educational disparity in smoking among adults had already been established back when they were in grade 7 (11). Children typically complete schooling decades before the onset of major illness. We see health inequalities for diseases in birth cohorts for which there was little if any useful health information available at the time they attended school.

Rather than just factual knowledge, schooling may form a set of enduring cognitive or emotional skills that foster health-promoting decisions throughout life. Literacy and numeracy are likely to help individuals make healthy decisions. More abstract skills, such as the ability to think abstractly, self-regulate, delay gratification, or adhere to organizational rules, may also be important. A large literature suggests that cognitive engagement itself—for example, that derived from intellectually demanding leisure time activities—may protect against dementia and promote

cognitive plasticity after neurologic injury (78). Education may promote lifelong participation in cognitively challenging activities, which may in turn increase the chances of better health and survival.

Time spent in school is also time *not spent* engaged in other activities, some of which may be health-damaging. Many of the endeavors crowded out by school-time may be unhealthy for adolescents, for example, drug and alcohol use, criminal behavior, sexual activity, or working in physically noxious situations. Historically, schooling of adolescents has competed with work, and that work has often been arduous and potentially unhealthy. For more recent cohorts, there is a spike in risky behavior and delinquency during hours immediately after school, suggesting that "warehousing" effects may be important (79).[12]

Finally, education may improve long-term health by increasing the chances that you will have a well-educated spouse, well-educated friends, and well-educated acquaintances. As discussed in Chapter 7, this social network may provide many health advantages. Indeed, a major benefit of attending Ivy League schools derives from the powerful social connections (the "social capital") that it provides.[13] There is a distinct advantage in rooming with a Crown Prince or a future Internet mogul in one's freshman dormitory, even if most of their time is spent partying and getting drunk. In short, education is a bundled process, not a binary "exposure" in the ordinary epidemiological sense.

RANDOMIZED TRIALS OF EARLY EDUCATION

Conducting randomized trials of elementary school education (deemed "compulsory" in most countries) is not ethically feasible; that is, one cannot flip a coin to force one group of children to stay in school while another group is tossed out. What we have is evidence in the form of randomized trials of early schooling, serving children younger than the typical school-starting age.[14]

Several interventions on schooling to either improve access to early education (e.g., at age 4) or improve quality of early schooling have been evaluated with randomized controlled designs. These studies have been very influential despite generally small sample sizes and limitations in the available health outcomes. Evaluations to date have emphasized cognitive, educational, and labor market outcomes, rather than health per se. Participants are relatively young for evaluations of many long term health effects, and thus our understanding of the health consequences of these interventions is quite preliminary. However, evidence from these studies is very encouraging about the likely causal effects of early childhood educational experiences on health.

12 Of course, if warehousing were the main explanation for the protective health effects of education, there would be much cheaper ways to promote health than building schools and investing in the training of teachers. Parents could achieve the same result by locking up their children in their basements.

13 As opposed to the quality of teaching in the classrooms. For example, Ivy League sophomores sometimes express frustration that their classes seem to be taught by teaching fellows rather than by their professors (*Harvard Magazine* "Life without Mr. Chips," July-August 1995. Accessed at: http://harvardmagazine.com/1995/07/life-without-mr-chips)

14 The reason why these trials were ethically permitted in the 1960s and 1970s was because they went beyond the prevailing standard of care, that is, children assigned to the control group were not missing out on "usual care."

The Perry Preschool Project randomly assigned disadvantaged African American children in Ypsilanti, Michigan, to an experimental group (n = 58) receiving center-based preschool, home visits, and parent group meetings or to a control group (n = 65). Perry Preschool has received a great deal of attention despite the relatively tiny size of the sample because the rigorous design provided compelling evidence that the program more than paid for itself. With follow-up through age 40, a cost-benefit analysis suggested a $12.90 return for every $1 in program cost—notwithstanding the intensive costs associated with setting up the program, for example, paying for preschool teachers who were qualified with at least a masters degree and who provided instruction to children for 2.5 hours per day for 30 weeks per year. Stated differently, the Perry Program was like a magic piggy bank that yielded $13 of "profit" for every dollar put in it. The financial returns to Perry Preschool derived largely from reduced criminal involvement of children receiving preschool benefits, but these analyses did not include any financial consequences of improved health among beneficiaries (80). In general, children randomized to preschool had better health and lower cumulative mortality (3.4% vs. 7.7%) 37 years later, at age 40, although these differences were not statistically significant. Combining multiple indicators, Muennig concluded there was evidence that overall health status (based on mortality, self-report of stopping work due to health, and self-reported health) was better for children randomized to preschool than for the comparison group, but self-report of health conditions, especially joint pain, was worse for those who attended preschool. He also found evidence of lower use of tertiary health care services and lower drug use (81).

The Abecedarian Study, conducted in the 1970s in Chapel Hill, North Carolina, randomized children to an intensive educational program provided in a year-round childcare facility from infancy to age 5. At age 5, before entering kindergarten, children in both Abecedarian treatment and control groups were rerandomized to control or to an experimental group that received additional services in kindergarten through grade 2. Thus, Abecedarian followed a factorial design. At age 21, Abecedarian children who received early childhood educational services had lower adolescent pregnancy rates, fewer depressive symptoms, and were less likely to smoke marijuana (82, 83). At age 30, Abecedarian children who had received preschool services had higher average years of education (13.46 vs. 12.31) and had older ages at first birth (21.78 vs. 19.95 years, $p = 0.03$), but no statistically significant differences in criminal conviction rates, internalizing or externalizing behaviors, self-rated health, binge drinking, or marijuana or cigarette use. However, testing a joint outcome of depressive symptoms, self-report of "no health problems since age 15," and no hospitalizations in the past year, the preschool group did have significantly better health. Similarly, although no individual health behavior was significantly better among individuals randomized to receive preschool services, a combined outcome of 11 behaviors (addressing automobile safety, drug, tobacco, and alcohol use, and having a primary care provider) was significantly better (84).

As discussed above, the Tennessee Project STAR experiment focused not on early education but on enhancing the quality of education experienced by randomizing children to spend kindergarten to third grade in one of three groups: small class size; regular class size with a classroom aide; or regular class size without an aide. The study initially randomized 328 kindergarten classrooms, including 6,325 children in 79 schools; an additional 5,456 new children enrolled in the schools over the following 4 years, and they were also randomized. Although children were originally randomized to one of three groups, the classroom aide group appeared very similar to

the regular classroom, so analyses typically contrast the small classroom group to the regular-sized classroom group, without regard to the presence of the classroom aide. Muennig et al. reported that mortality through age 29 was unexpectedly *higher* among children assigned to small class sizes (85). Puzzlingly, the lower mortality rate emerged within only a few years of randomization, that is, before age 10, and persisted roughly through the first 12 years of follow-up (through age 17), after which survival gaps began to narrow. Changes in the hazard ratio over age may suggest different effects of educational quality on the major causes of death in early life versus chronic conditions that emerge in adulthood. Effect estimates were very imprecise due to the small number of deaths (n = 146) in this young sample, so follow-up is critical. The only other health outcome evaluated for the STAR Project is disability status, indicated by receipt of Social Security disability benefits among participants who could be matched to Social Security records; there was no difference in this outcome in this subgroup (86).

In summary, the randomized trials show consistent benefits of early education and educational enhancements on *educational* outcomes, and to some extent on labor market outcomes such as employment and earnings. The evidence for noncognitive health benefits is sparse but suggestive. Convincing results await larger studies, longer follow-ups, and more comprehensive health assessments.

HEAD START STUDIES

Head Start is a US federal program intended to improve the learning skills, social skills, and health status of poor children so that they can begin schooling in a condition equal to their more socioeconomically advantaged peers, thus providing a "head start." The program provides preschool, health, and other services to poor children aged 3 to 5 and their families. It was established in 1964 as part of the "War on Poverty," and has enjoyed great public support. In fiscal year 2012, Head Start served nearly a million children, with a federal budget of around US$8 billion. Like the Perry Preschool and Abecedarian Programs, Head Start is motivated by the notion that human capital investments in the early years of life may be crucial for the development of cognitive and noncognitive skills, and potentially more important than interventions in later years.

Unlike the Perry Preschool or Abecedarian Programs, Head Start was not implemented following a randomized design. Recent randomized evaluations of Head Start have reported significant but modest impacts on cognitive test scores shortly after implementation, which seem to fade away quickly after completion of the program (87, 88). However, caution is warranted in interpreting these evaluations because they do not tell us about the long-run effects of Head Start; the jury must wait years or decades to obtain credible estimates of the effects of the program on health in the long term (89). Current evidence on the long-run health effects of Head Start relies primarily on nonrandomized evaluations following quasi-experimental designs. Evidence from Head Start offers an important complement to earlier randomized evaluations because it addresses the question of how preschool education programs implemented "in the real world" can impact outcomes.

The most important evidence from Head Start stems from within-family comparisons of siblings comparing children participating in the program with siblings in the same family who did not take part in the program. Sibling designs control for all shared family characteristics that may

influence the likelihood of participating in the program. Conversely, they rely on the assumption that within families, parents do not assign different children to preschool versus home or other daycare services on the basis of the child's characteristics. Based on this design, Garces and colleagues (90) used data from the Panel Study of Income Dynamics to assess the long-term effects of Head Start, focusing primarily on children who attended in the 1970s. They found significant educational benefits and reduced criminal involvement among African American children who participated in Head Start. Following a similar approach, Deming assessed the long-term effects of Head Start for children who attended between 1984 and 1990. His analyses are based on the children of mothers in the National Longitudinal Survey of Youth, comparing siblings who differed with respect to their participation in the program, and controlling for an extensive set of confounders (91). He finds that when children were 19 years or older, those who participated in Head Start had significantly better educational outcomes and self-reported health, but there were no significant differences in criminal involvement or teen childbearing. Using a similar approach, other studies have found that participation in Head Start during early life is associated with lower probability of smoking cigarettes in young adulthood (92).

The implementation of Head Start bundled educational services with health screenings and immunizations, nutritional supplements, and other services. Most of the evidence from Head Start focuses on cognitive outcomes, but more recently studies have started to assess impacts on health. Ludwig and colleagues (93) exploited a discontinuity in program funding across counties that resulted from program implementation. In the midst of President Johnson's War on Poverty (specifically during the spring of 1965), the Office of Economic Opportunity (OEO) provided technical assistance to the 300 poorest counties to develop Head Start proposals. Eligibility for the program was based on an arbitrary cut-off which classified some countries as "poor" or "nonpoor" and led to larger participation rates and more funding to "poor" counties just below the cut-off for eligibility (the "treatment" group) as compared to "nonpoor" counties just above the eligibility cut-off (the "control" group). The assumption behind this design is that outcomes are smooth around the cut-off, because the threshold was defined on an arbitrary criterion a few years before the implementation of the program. As a result, counties just below the cut-off are comparable to counties just above the cut-off. This analysis found marked reductions in child mortality from ages 5 to 9 from causes plausibly associated with Head Start, an effect large enough to drive mortality rates from these causes in the treatment counties down to the national average for extremely disadvantaged children. There were no effects on causes of death believed to be unrelated to program exposure.

While the causal impact of Head Start programs continues to be debated on a number of outcomes (94), there is much agreement that the program, when adequately delivered, has important cognitive and social benefits. These changes are likely to lead to better health in the long term, even though to date there are few informative analyses of adult health (in part because the earliest Head Start beneficiaries are only now reaching middle age). This is a very promising area for future research, not only to elucidate the long-term health effects of Head Start but also to better understand the mediators of such effects, heterogeneity in the magnitude of effects across different children, and how the context modifies the effects. Criminal involvement, so important for the cost-benefit analyses of Perry Preschool, may be more relevant for cohorts who reached adolescence at the height of the US crime wave, but less relevant for children growing up in lower-crime areas or time periods.

INSTRUMENTAL VARIABLE ANALYSIS OF COMPULSORY SCHOOLING

Legal regulations regarding the number of years of school a child must attend changed rapidly in many western countries in the twentieth century. For example, the mandatory age of school enrollment dropped in many US states from ages 7 or 8 at the beginning of the century to age 6 by World War II. In addition, the earliest age allowed to drop out of school or obtain a work permit was raised in most states, often from age 12 or 14 to age 16 or 18. As discussed above, these increases in mandated schooling appear to have increased average years of education attainment at least modestly. The effects in the United States were generally small because the laws were not consistently enforced and, by the time new laws were introduced, most students were attending more than the minimum requirement anyway. Lleras-Muney showed that the laws had very little effect on African American children, likely due to lack of enforcement (95). Outside the United States, changes in mandatory schooling often had larger effects. In the United States and many European countries, these legal changes have been treated as "natural experiments" to estimate the health effects of slight increases in educational attainments. These natural experiments are compelling because individual preferences, talents, and health conditions—generally considered potentially serious confounders in research on the health effects of education—presumably have little effect on mandatory schooling laws. If changes in mandatory schooling laws predict differences in population health, it suggests increases in education did induce better health.

In fact, results of compulsory schooling law (CSL) natural experiments have been mixed, with evidence of important benefits for cognitive outcomes, even in old age (96–98), but contradictory results for other domains of health (99–101). The most common limitation of CSL studies has been imprecise effect estimates. Even in countries in which the legal changes affected a large fraction of children (as was often the case in Europe), the change typically only induced a single year of additional education, which we would anticipate to have quite small effects. For example, Silles uses CSL changes in England and Ireland to estimate the effect of education of an individual on the health of that person's children (102). This is a very substantively important research question, and the conventional analysis, using ordinary least squares (OLS), indicates a highly statistically significant beneficial effect. If this effect is causal, it bodes well for the health of future generations, given recent rapid increases in parental education. The IV effect estimate had a point estimate associated with harmful effect of education, but the standard error was so large that the confidence interval also included a plausible magnitude of (beneficial) effect 5 times larger than the OLS estimate. In short, the IV estimate was consistent with both huge beneficial and huge harmful effects of education. This analysis included 100,928 individuals. The most precise estimates are from huge surveillance data sets and thus focus on outcomes such as mortality. Findings from several countries suggest very small or null effects of an extra year of mandated schooling on mortality (100, 101, 103), but a study across 12 countries (104) and a careful examination of the implementation of reforms in Denmark (105) suggest that the estimates differ substantially based on the magnitude of the education reform (e.g., 1 year versus 2 or more years of additional schooling), the implementation of the policy, and across individual characteristics, such as gender, parental SES, and cognitive ability.

Findings from CSL reforms at this point are a mixed bag, but this line of research is likely to be very informative in the long run. However, it is important to recognize what natural experiments based on CSLs can and cannot tell us. Prior research cannot, for example, tell us about the likely effects of college completion, because college completion was not mandated in any of the countries studied. Nor does the research address the consequences of primary school access; nearly all studies to date focus on increases in education in the range of 7 to 12 years. These studies cannot tell us about the effects of schooling on children who would have pursued schooling even if not legally mandated to do so, that is, high-achieving students or students for whom the barriers to schooling were about access rather than interest.[15] Finally, because CSLs impact *everyone* in a cohort, they cannot estimate any effects of education that are via *relative* position or status. Compulsory schooling laws may not modify anyone's relative position in society, or may even reduce the relative value of credentials as such credentials become more common.

Despite these limitations, the inconsistent evidence on health effects of increases in CSLs is troubling. Could CSLs have had beneficial effects for some people in the population but harmful effects for others? Consider the results from the Tennessee STAR classroom quality analyses described above (68). For children who would have attended more than the mandated schooling minimum regardless of the law, increases in CSL may have implied larger class sizes, less experienced teachers, and more classmates from disadvantaged backgrounds. All of these changes may have harmed the children who would have attended school regardless of the legal mandate. The implication of this is that we should worry tremendously about the gap between the measure of education as we typically employ it in regression models (years of schooling completed) and education as experienced by children (in a classroom with or without appropriate resources, with a good teacher or a bad teacher, with strong classmates or weak classmates). Increases in educational mandates may not benefit children's health unless they are implemented in a way that maintains quality standards for all children.

Promising future initiatives related to the CSL analyses will leverage other experimental or quasi-experimental education projects. In the last decade, there have been tremendous changes in the structure of schooling in the United States, including the charter school movement (often implemented with lotteries), increasing access to early childhood education, and vast expansions in postsecondary education. In the United States, increases in college attendance were especially marked among women, blacks, and Hispanics (106), so the possible impacts on social inequalities could be large. Rigorous evaluations of health effects of these initiatives will help to prioritize effective educational policies and to understand how SES affects health.

INTERPRETING TRENDS IN EDUCATIONAL INEQUALITIES

Recent evidence indicates that educational inequalities in mortality are growing larger in the United States (107, 108). However, trends in social inequalities in health over time do not necessarily

15 In other words, IV analyses can only tell us about the effect of the exposure within the range wiggled by the instrument, the so-called local average treatment effect.

imply deterioration (or improvement) of conditions for the disadvantaged. Paradoxically, social inequalities in health may *increase* over time if some individuals are moved out of the most disadvantaged social categories, for example, if population-average education is increased. The UK Black Report was the first to entertain this possibility, namely, that upward social mobility would leave behind a (shrinking) pool of the most disadvantaged people, resulting in apparent widening of health inequalities (44).

To understand trends in health inequalities across *modifiable* social characteristics such as education, it is thus important to evaluate the role of changing "composition" of low-education versus high-education groups. Many factors presumably influence attained education, including parental education, cognitive skills, and health. The average background characteristics of low- versus high-education individuals may change over time. Thus, consider the findings of Montez and Zajacova: Age-standardized death rates for US white women ages 45–84 with 0–11 years of education increased by 21% between 1986 and 2006 (from 0.0235 to 0.0284), while mortality rates for college-educated women declined by 11% (from 0.0066 to 0.0059) (108), leading to an increase in the rate ratio from 3.55 to 4.82 over the 20-year period.

Two alternative interpretations of such trends in educational inequalities are possible. One possibility is that the declining life expectancy of low-education women reveals a modern problem: The circumstances shaping health for low-education women have deteriorated over the past 20 years, leading to increased mortality. This implies, for example, that a 45-year-old woman with low educational attainment in 2006 would have had a higher risk of mortality than the same woman had she been age 45 in 1986. This is a devastating result because it suggests that women in the lowest educational stratum have not benefited from the medical and social advances of the last two decades.

An alternative interpretation of secular trends in educational inequalities in health is (somewhat) more benign. The fraction of women with less than high school education was much smaller in 2006 than in 1986 and the women with less than high school education in 2006 were likely more disadvantaged in other regards—such as family and social history, even before attending school—than women with comparable education levels in 1986. For example, in the 1940 birth cohort, it was not unusual for a woman from a middle-income family to leave education before completing high school. Such women had various advantages to offset low education, and high school dropout was not an indicator of special cognitive or social difficulties. However, by the 1960 birth cohort, it was quite unusual for middle-income women to drop out of high school. In general, women from the 1960 birth cohort who did not complete high school came from very disadvantaged circumstances or had other major educational challenges. The between-birth-cohort changes in "selection" into the low-education group were especially acute for white women. Thus, the changing life expectancy for low-education US women may reflect simply that lower-educated women included an ever more selected group of disadvantaged women over time, rather than implying that the environments encountered by disadvantaged women are becoming more harmful. Empirical evidence to distinguish these two alternative explanations would be based on comparing the characteristics of the least educated to the most educated women in 1986 versus 2006, but would necessarily use only characteristics that could not themselves be affected by education.

INCOME AND HEALTH

We now turn to the relation between money and health. Evidence from studies across many different settings and populations suggests that more income and wealth are associated with better health (109–113). The causal inference challenges relevant to education and health are also a concern for interpreting the association between income and health, although the most plausible specific mechanisms generating biases differ. Social epidemiologists often interpret this association as reflecting causality from income to health: Higher income may enable better access to the means to produce good health, including better access to health care, as well as other forms of "healthy consumption" such as better housing, means of transportation, or clothing. On the other hand, this relationship may arise from the fact that poor health impinges on the ability to work, thereby reducing income and wealth accumulation (114). As previously discussed, the relationship between income and health may also arise from unobserved or unmeasured characteristics that are correlated with both income and health, such as early-life investments, parental SES, or lifetime preferences.

Despite the overwhelming evidence that income is correlated with health, there is a paucity of evidence on whether income is causally related to health. For example, a recent systematic review on the causal impact of income on children's outcomes (including health) identified more than 46,000 articles that matched the search criteria, but only 34 studies were judged to meet the inclusion criteria, which included an experimental or quasi-experimental design. In this section, we review a selection of studies using experimental or quasi-experimental designs to examine whether income has a causal impact on health. In terms of methodology, studies addressing this question can be classified in two broad categories: The first category includes randomized experiments whereby health in a group of individuals randomly allocated to receive some form of income transfers (treatment group) is compared to health in a group of individuals randomly allocated to receive no income transfer (control group). The second and most common type of evidence comes from natural experiments, which exploit the fact that income is sometimes assigned to individuals or groups exogenously or "as-if-random" due to conditions determined by nature or other factors such as a change in benefit entitlement laws, lotteries, or unexpected stock market gains.

INCOME FROM WELFARE-TO-WORK PROGRAMS: EXPERIMENTAL STUDIES

Experimental evidence of the impact of income on health comes from the evaluation of income support and other welfare programs often targeted to poor households. The advantage of these studies is that exposure to income is based on random assignment. In addition, these studies directly assess whether a policy-induced change in income improves health, enabling us to anticipate health changes we may expect through public policies that transfer income. Importantly, income changes induced by policy may be fundamentally different from changes in income induced by temporary income shocks such as lottery winnings. On the other hand, a disadvantage

of these studies is that government income transfers are rarely isolated entitlements. They are often accompanied by requirements to work or engage in education, or in-kind benefits for training or childcare. Evaluations of policies that induce income changes are thus not always clean quasi-experiments of the impact of additional income on health, but may encompass the impact of a "policy package," with income changes being only one component. Nevertheless, with some creativity, quasi-experimental studies can sometimes come close to isolating income effects from the effect of nonfinancial components of social programs.

A substantial part of the literature on the impact of welfare programs affecting income has focused on children. Emphasis on the early years of life in recent public policy debates reflects the growing view that human capital formation starts very early in life (115–119). In a "household production" model, child outcomes are conceived as the product of the amount and quality of parental time inputs, the amount and quality of other caretakers' time, and market goods spent on behalf of children (120–122). Income is important because it enables parents to purchase inputs that matter for the production of positive child outcomes. Exposure to income shocks in the critical years of a child's development may thus have both short- as well as long-run effects on health.

Experimental evidence on the impact of income in the early years comes primarily from welfare-to-work experiments conducted in the 1990s. These programs randomly assigned low-income and welfare-recipient single parents to a variety of welfare and employment policy treatments, or to a control group that continued to receive welfare as usual. In a recent study, Duncan and Morris (120) used data for 16 of these programs to estimate whether variation in income that arises from random assignment to the treatment group has an impact on children's developmental outcomes. All of these programs sought to increase the self-sufficiency of low-income parents. However, some programs focused exclusively on increasing employment and reducing welfare use, while other programs also increased parental income through generous earning supplements.

Using treatment assignment as an instrumental variable, Duncan and Morris assessed the impact of income changes induced by these programs on child outcomes. To separate the impact of income supplements from that of childcare and educational subsidies, they only incorporated a selection of programs that exclusively offered income supplements. To separate the effect of income from that of increased employment, they included the latter as a control variable, thus estimating the impact of income net of employment effects. Results from this study suggest that income has strong positive effects on children's cognitive outcomes and educational achievement. In particular, a $1,000-increase in income was associated with between 0.06- and 0.60-standard deviation increases in child achievement, as measured by a combined measure of both cognitive scores and parental reports. Comparable effects on children's developmental outcomes have been reported in evaluations of similar welfare programs (123–125). Interestingly, these effects appear to be confined to children in the very early stages of development, from age 0 to 5, with less consistent or no effects for children older than 5 years (125). This provides some evidence of the potential benefits of income in the critical years of early childhood development, where the family environment is potentially more important than the school or childcare environment.

INCOME FROM WELFARE-TO-WORK PROGRAMS: NATURAL EXPERIMENTS

Alongside evidence from experimental studies, a recent stream of literature on the impact of policy-induced changes in income has focused on the health effects of the Earned Income Tax Credit (EITC), the largest US federal antipoverty program for the nonelderly, established by Congress in 1975 and expanded during the 1990s. The EITC is a refundable credit originally designed to encourage work by offsetting the impact of federal taxes on low-income families. The program has been shown to increase employment, particularly among single mothers; some studies suggest that EITC may be responsible for about two-thirds of the rise in labor force participation by single mothers between 1984 and 1996 (126). The boost in income for low-income families in the EITC program is substantial: On average, EITC provides a minimum-wage worker with two children an increase of around 40% in annual earnings.

The general conclusion from studies evaluating the health impact of EITC programs is that they appear to improve health outcomes among young children and some adults. However, some studies have reported detrimental health effects for some subpopulations, raising caution on the idea that income provided through welfare-to-work programs are universally beneficial for health. A potential limitation is that studies often combine the effect of both cash transfers and work incentives, making it difficult to disentangle whether income alone has an impact on health. Because participation in the EITC program is not randomly allocated, evaluations of EITC use quasi-experimental designs that exploit aspects of program implementation that are uncorrelated with household's characteristics. In a recent study, Dahl and Lochner (127) exploited the large, nonlinear changes in the Earned Income Tax Credit expansion in the late 1980s and 1990s as an exogenous source of variation to identify the impact of family income on child achievement. Their design exploited the fact that the maximum benefit amount increased substantially over time, and the range of family income eligible for the ETIC also expanded, particularly benefiting low- to middle-income families. Based on panel data from children in the National Longitudinal Study of Youth (NLSY), they used these changes as instrumental variables to identify the impact of family income on child outcomes. Their results suggest that EITC significantly improves child cognitive outcomes. Instrumental variable estimates indicate that a $1,000-increase in family income raises math and reading test scores by about 6% of a standard deviation, with larger effects for children from more disadvantaged backgrounds, younger children, and for boys.

In a separate study, Strully et al. (50) exploited state variations in the introduction of the EITC program to examine whether induced increases in income and employment improved maternal and childbirth outcomes. They used a difference-in-difference approach, whereby changes in child and maternal outcomes in states that introduced EITC before and after enactment were compared with changes in outcomes in states that had not implemented the EITC program. Their results suggest that EITC increases birth weight and reduces maternal smoking during pregnancy. This beneficial effect was limited to mothers aged 19 to 34 years, while the program appeared to have no effects on younger mothers and led to increased smoking among mothers aged 35 years and older. However, recent studies using different identification strategies suggest that EITC decreased the probability of smoking for mothers who benefited from

an expansion of income through the program, relative to those who did not benefit from such expansion (128, 129).

A word of caution is required with regard to the impact of EITC on some adult outcomes. A recent study (130) used the generosity in maximum state and federal EITC benefits that a family could receive as an instrumental variable to assess the effect of family income on health among working-age adults. Results from this study suggest that family income has no consistent effect on self-rated health or the prevalence of most functional limitations in the next year. Using a similar identification strategy, a separate study (131) found that income from the EITC significantly raises BMI and obesity in women with EITC-eligible earnings. This effect is not trivial; simulations suggest that the increase in real family income from 1990 to 2002 as a result of the EITC program may account for between 10% and 21% of the increase in women's BMI and 23% to 29% of the increase in women's obesity prevalence.

In conclusion, it would seem that EITC and similar in-work tax credits improve birth and developmental outcomes among children. However, in line with results from a recent systematic review (132), the evidence on adults is mixed and insufficient to draw firm conclusions, with contradictory evidence regarding the impact of EITC on obesity and adult health outcomes.

CONDITIONAL CASH TRANSFERS

Conditional cash transfers (CCTs) are programs that transfer cash to poor households on the condition that they comply with a set of prespecified behavioral requirements to invest in the human capital of their children (133). On health and nutrition, conditions often require periodic attendance at medical check-ups; growth monitoring and vaccinations for children less than 5 years old; perinatal care for mothers; and attendance at periodic health information sessions. In terms of education, CCT programs require that children enroll in school and attend 80–85% of school days; and in some cases, they require some level of school performance. Most CCT programs transfer payments to the mother in the household, under the assumption that mothers are more likely to invest resources in the human capital of their children.

There are two sets of arguments for attaching conditions to cash transfers (133). The first set of arguments is based on the notion that parents underinvest in the human capital of their children. Parents may hold misguided beliefs about the process and returns of investments in their children, and they may therefore "underestimate" the potential benefit of sending their children to school or having regular medical check-ups. The second argument refers to the political economy conditions necessary to implement these programs using public funds. Redistribution to the poor through cash transfers may enjoy little societal support unless it is perceived to be conditioned on "good behavior." Attaching cash transfers to conditions focusing on building human capital among children may add to the political acceptability of CCT programs, which may otherwise be perceived as paternalistic, or enjoy little support from the middle classes, as they do not directly benefit from the program.

Over the last 15 years, CCT programs have become increasingly popular. They were initially developed in Mexico through the Oportunidades program, and today almost every Latin American country has a CCT program. Large-scale CCT programs have also been implemented in other countries such as India, Bangladesh, Indonesia, Cambodia, Malawi, Morocco,

Pakistan, South Africa, and Turkey. Programs have also become substantially larger over the years, serving millions of families in the developing world. Today, the Oportunidades program in Mexico serves over 5 million households, while the CCT program in Brazil, Bolsa Familia, serves 11 million families (46 million people). Conditional cash transfer programs are thus the largest social programs in many countries in Latin America. They have been promoted as a way to reduce poverty and increase human capital formation, helping households to break the cycle of poverty transmission from one generation to another (133). Interest in CCT programs has even spread to high-income countries—most recently, New York and Washington, DC, have implemented CCT programs as a means to improve school attendance among children of low-income households.

What do CCTs tell us about the causal impact of income on health? A unique feature of many CCT programs is that they have been accompanied by careful evaluations, often using a randomized design. When randomization did not proceed as initially planned, studies have used alternative quasi-experimental methods such as regression discontinuity design. A downside of CCT programs—in terms of what they can tell us about how income influences health—is that they are cash transfers with strings attached, in the form of behavioral requirements. As such, the impact of income from CCT programs does not directly correspond to the structural parameter for the effect of a pure income shock on health. Nonetheless, they offer a powerful evidence base of what income changes induced by a policy that alters income through conditional transfers can achieve in terms of health.

The story that CCT program evaluations tell us is in some ways similar to that emerging from evaluations of the EITC or other in-work tax credits. Essentially, there is evidence suggesting that conditional cash transfers have strong beneficial effects on child health, and in some cases, they may also improve mother's health. However, CCT programs may also sometimes lead to undesirable health outcomes. Most studies suggest that CCT programs increase the use of preventive health services. Because CCT programs most influence those who are less likely to use health services in the absence of the program, CCTs have contributed to substantial reductions in preexisting disparities in education and health. A less clear pattern emerges from studies examining the impact of CCT on morbidity outcomes. Conditional cash transfer programs seem to increase child height among some population groups, but not all studies have found these effects. There is also some evidence that CCT programs improve morbidity outcomes, including the risk of overall illness, diarrhea, and respiratory infections, although some evaluations have found no effect on these outcomes. Conditional cash transfer programs have been shown to increase food expenditures as share of total expenditures, and in some cases, they may also improve the quality of diet among adults and children.

Perhaps the most convincing nonexperimental evidence that CCT programs may be beneficial to health comes from recent studies evaluating their impact on infant mortality. Rasella and colleagues (134) used data for the period 2004–2009 across different municipalities to assess the impact of the Brazilian CCT program on infant mortality. Using a municipality fixed effect model, they exploited variations across municipalities in the level of coverage of the program over the study period. Their results suggest that moderate or high levels of coverage of the program significantly reduced mortality, in particular for deaths attributable to poverty-related causes such as malnutrition and diarrhea. A separate study found similar results and suggested that most benefits come from decreases in postneonatal mortality (135).

While the vast majority of studies suggest that cash transfers have beneficial effects on health, a cautionary note is needed on the potential short-term impact of CCTs on some behavioral outcomes. Based on data from the CCT program in Colombia, a recent study (136) showed that CCTs are associated with increased body mass index (BMI) and higher odds of obesity among poor women. A separate study (137) examined the causal impact of Oportunidades, the CCT program in Mexico, on overweight and obesity of adolescents in poor rural areas. Exploiting the fact that individuals were assigned to treatment based on a poverty score, the authors used a regression discontinuity design that took advantage of the discontinuity in program participation induced by the eligibility cut-off. The assumption is that individuals with a score just below the eligibility cut-off are comparable to those just above. Results suggest that CCTs led to a decrease in obesity among adolescent women. However, CCTs also led to higher rates of smoking initiation among adolescent females. A more detailed evaluation of the dietary effects of CCTs in poor rural communities in Mexico found that transfers increased household fruit, vegetable, and micronutrient consumption, but also led to excess energy consumption (138). Overall, these results suggest that income from CCT programs may have many health benefits, but in some cases they may also lead to higher consumption of unhealthy goods.

An important consideration is that current evaluations have only been able to assess short-term effects of cash transfers on health, and they may therefore underestimate the potential long-run effects of human capital investments arising from CCT program exposure during early life. Virtually all program evaluations have found that CCTs increase school enrollment, particularly for those who have low enrollment rates to start with. These effects can be very substantial; the CCT program in Nicaragua, for example, increased school enrollment by 12.8%, from a baseline of 72%; Chile Solidario, the CCT program in Chile, increased school enrollment by 7.5%, from a baseline of 60.7% (133). If these increases in school enrollment lead to higher educational achievement, conditional cash transfers in the critical early years may have important returns in terms of career and earning trajectories, ultimately leading to better health outcomes in later life.

LOTTERIES, INHERITANCES, STOCK MARKET, AND LOCAL ECONOMY SHOCKS

While policy-induced changes in income are perhaps most useful from a public health standpoint, natural experiments may also exploit unanticipated changes arising from other shocks to income or wealth unrelated to policy. Following this approach, studies have used lotteries, inheritances, stock or house price fluctuations, as well as major changes in the local economy, to assess the impact of income on health. The assumption is that income shocks arising from these sources are assigned "as-if-random," mimicking a randomized trial. Generally, many of these studies have found relatively weak evidence of an effect of income or wealth on health. In a seminal paper, Smith (17) used data from the Prospective Study of Income Dynamics (PSID) to assess whether the largely unanticipated wealth gains from the stock market boom in the late 1980s and 1990s in the United States led to short- and medium-term gains in health. The assumption is that this surge captures unanticipated exogenous wealth increases unrelated to a person's health. Whether looking at a short frame of five years or a longer horizon of a decade or more, his findings suggest

that changes in stock wealth do not influence the onset of future health conditions or changes in self-reported measures of health.

Lindahl (139) examined the impact of positive income shocks arising from lottery prizes on health and mortality in Sweden. Because a lottery randomly draws winners from a pool of participants, lottery prices generate exogenous variation in income among lottery players. Findings suggest that higher income is causally associated with better health, so that a 10% increase in income improves self-rated health by approximately 4–5% of a standard deviation, and reduces the risk of dying over the next five to ten years by 2–3 percentage points. Following a similar approach using British Panel data, another study (140) found that exogenous income shocks from lottery winnings were associated with more smoking and social drinking, but they had no effect on physical or mental health measures.

A series of studies have used inheritances to examine the impact of exogenous wealth shocks on health. Although inheritances are not randomly distributed, the timing of inheritances is likely unrelated to a person's health, as they often come as unanticipated shocks to wealth. Studies on the impact of inheritances often introduce *individual fixed effects*, which essentially assess whether within-individual changes in wealth as a result of an inheritance are associated with individual-level changes in health. Fixed effects control for all variables that are constant across individuals, for example, race, gender, parental investments in early life, parental SES, and educational attainment, although they do not control for time-varying confounders, which may be correlated with both changes in income and health. In the case of inheritances, this may not be a major concern if the timing of inheritances is largely unanticipated; on the other hand, if inheritances are concomitant with the death of a parent or spouse, estimates of effect will combine the impact of inheritances with that of bereavement and associated changes.

Using data from the Panel Study of Income Dynamics (PSID), Meer et al. (141) found that changes in wealth as a result of a recent inheritance had no effect on self-rated health. Michaud and van Soest (142) used longitudinal data from US Health and Retirement Survey (HRS) participants aged 51 to 61 and found no evidence that changes in wealth induced by inheritances have any impact on a variety of health outcomes including several medical conditions, physical functioning, and depression scores. Some of these studies instead suggest that the negative impact of poor health on labor supply and the ability to accumulate financial resources are important drivers of the association between health and wealth (142) or income (17). More recently, Kim and Ruhm (143) used HRS data to examine whether inheritances affect mortality rates, health status, and health behaviors among older adults. Their findings suggest that inheritances increase out-of-pocket expenses on healthcare, utilization of medical services and light drinking, and that they reduce obesity. However, they conclude that inheritances have no substantial effect on mortality.

Findings from inheritance and lottery studies would seem to suggest that income shocks do not consistently influence health. On the other hand, these studies focus on transitory income shocks, and may offer little insights into how permanent income influences health in the long run. This is related to the fact that income from transient shocks is used for consumption in a different way than is permanent income. For example, a recent study (144) found that income from lottery winnings had no effect on most components of household's expenditures, including food at home, transportation, or total monthly outlays. Instead, the winner's consumption is largely confined to expenditures on cars and other durables. Temporary income shocks may

thus be used for the consumption of goods with little or no effect on health, as opposed to other income shocks such as CCTs, which influence monthly consumption and potentially increase human capital investments.

A natural experimental study that examined the impact of the introduction of a casino by a tribal government in North Carolina offers a somewhat different picture. This tribal government distributed approximately $6,000 per person to all adult tribal members each year. These income shocks come as regular payments and thus resemble more closely than lotteries or inheritances the nature of policy-induced income transfers. The study compared Native American children with non-Native American children, before and after the casino opened, and found that receipt of casino payments increased the educational attainment of poor Native American children in their young adulthood, and reduced criminal behavior and drug use. Costello and colleagues (145, 146) found an improvement in mental health outcomes in a North Carolina cohort of Native American children following the opening of a nearby tribal casino, a positive benefit that persisted into adulthood. Interestingly, the effect of extra income on children's BMI differed by household's initial SES. Children from the initially poor households had an increase in BMI as a result of the increase in income, while BMI decreased among children from wealthier households (147). Using a difference-in-difference approach, Wolfe and colleagues (148) found that income from the legalization of Native American casino gaming in the late 1980s had a positive effect on several indicators of Native American health, health-related behaviors, and access to healthcare. In contrast, a separate study (149) examined monthly mortality from accidental deaths among Cherokee Indians over 204 months from 1990 to 2006 and found that accidental deaths rose above expected levels during months of large casino payments, suggesting a detrimental effect of these income shocks in the short run.

CONCLUDING REMARKS ON THE CAUSAL IMPACT OF INCOME ON HEALTH

The studies above reflect the complexity of isolating the causal effect of income on health due to reverse causality and factors that jointly determine income and health. Recent studies address these concerns by exploiting exogenous changes in income generated by unanticipated shocks such as inheritances or lotteries, as well as policy-induced changes in income resulting from in-work credit or conditional cash transfer programs. The assumption is that these changes in income are assigned "as-if-random," and therefore offer an opportunity to assess the causal impact of income on health.

Overall, many studies suggest that the association does not always reflect causality from income to health, and that social epidemiologists have often underestimated the reverse pathway. Nonetheless, in many circumstances, income does appear to have a causal beneficial impact on health, but the effect depends on the population studied, whether income shocks are temporary or permanent, and the period in the lifecourse in which income changes are experienced. While findings are sometimes contradictory, the evidence would appear to suggest that policy-induced income changes through conditional cash transfers that require human capital investments have substantial positive effects on the health of children, and often also on the health of mothers.

Income shocks in the very early years of life, from age 0 to 5, appear to have larger effects on health than income shocks at later stages of the lifecourse. On the other hand, temporary income shocks, such as lottery winnings or inheritances, do not appear to consistently improve health, and may in some instances promote health-damaging behavior. Similarly, in the short run, income from conditional cash transfers and in-work credits may sometimes increase BMI and cigarette consumption. However, in the long run, these and other permanent changes in income such as those arising from the introduction of casino gaming by tribal governments appear to bring important benefits to the health of poor households.

There is insufficient evidence on the long-run impact of policy-induced changes in income on health, as income transfer programs have only recently been introduced. However, existing evidence suggests that conditional cash transfers and in-work credits might have important effects on child developmental outcomes and school enrollment, potentially signaling a pathway toward long-run improvements in health for beneficiaries. The key challenge for future research is to estimate to what extent these programs lead to sustainable improvements in health by reducing poverty and improving the educational, labor market, and social trajectories of children in poor households.

EXCEPTIONS: WHEN IS HIGH SES NOT BENEFICIAL?

Income may enable healthy consumption but, as noted above in the discussion of lotteries and CCT programs, income may also facilitate access to unhealthy consumption of goods such as cigarettes, alcohol, or unhealthy foods. Economic theory is useful in helping us to understand why higher income from wages in the short run may sometimes lead to lower investments in one's health due to increased "opportunity costs," that is, in a situation in which an individual has to choose among several alternatives, the opportunity cost refers to the loss from not choosing the other alternatives. Individuals confront this kind of trade-off between choosing to spend more time working versus investing in their health. Suppose you earn a wage of $10 an hour and you decide to consume an extra hour of "leisure" (e.g., by going to the gym or cooking at home instead of eating in a fast food restaurant). How much does it cost you to take this extra hour of leisure? The extra hour of leisure effectively costs you $10 in forgone wages. Now suppose your employer decides to raise your wage to $13 an hour. It will now cost you $13 in forgone wages to go to the gym or cook at home for an extra hour. The implication is that higher wages in the short term may be associated with greater opportunity costs, which according to microeconomic theory would reduce the amount of time devoted to health-promoting leisure activities (114). However, the long-term pattern is the opposite—richer people tend to exercise more than poor people. How can these observations be reconciled?

To cite an additional example, the evidence on how business cycles affect an individual's health behaviors is surprisingly mixed. For example, long-term unemployment is associated with higher smoking prevalence (150–152). Yet there is also evidence that declining personal income among young people is associated with *decreased* odds of smoking, a finding that is consistent with the income elasticity of demand (the less one's income, the less one can consume) (153). Xu examined the effects of wages and working hours on health behaviors of individuals with low

levels ofeducation, using variations in wages and hours caused by changes in local economic activity (154). The analysis utilized a two-sample IV approach to combine data on individual health behaviors from the Behavioral Risk Factor Surveillance System (BRFSS) and the National Health Interview Survey (NHIS) with data on individual employment from the Current Population Survey (CPS). The authors found that rising wages associated with economic expansions are linked to greater consumption of cigarettes. The study also found that increases in the hours of work caused by economic expansions are associated with more cigarette consumption, less physical activity, and fewer physician visits. Interestingly, the evidence suggested that the effects of working hours on health behavior stemmed primarily from changes at the *extensive margin* of employment (i.e., changes in employment status), rather than from changes at the *intensive* margin (i.e., changes in hours of work conditional on being employed) (154). These findings imply that changes in earnings (and the attendant changes in the opportunity costs of time) may have heterogeneous impacts on time-intensive activities (e.g., exercising) versus less time-intensive behaviors (such as smoking) (154). In summary, the next generation of studies in social epidemiology calls for a more nuanced understanding of health production that recognizes the heterogeneous effects of income on specific health behaviors and outcomes and the distinction between long-term and short-term effects.

It is also worth considering situations in which health outcomes among the less educated are as good as or better than those among the well educated (155). To the extent that higher income facilitates unhealthy behaviors, higher education may also have similar consequences through its effect on income. As with income, education may be harmful when access to a highly desirable resource is extremely unhealthy. For example, in many contexts in sub-Saharan Africa, HIV infection rates are higher among more highly educated individuals (15); although education is associated with more consistent condom use, it is also associated with sex outside of marriage (156). Importantly, the association between education and HIV may change as a country transitions through stages of the HIV epidemic.

Education may not be advantageous when prevailing knowledge about a particular health condition or health practice is incorrect (e.g., that river blindness is caused by sorcery; or that taking a multivitamin pill can prolong life). Education is advantageous when combined with correct knowledge about the etiology of disease. For example Preston and Haines (157) describe how the children of physicians in late nineteenth-century New York were no less likely to die than children in the general population. However, following the acceptance of germ theory, physicians were among the first to benefit from the application of its principles (hand-washing and other hygiene practices), resulting in a sharp drop in mortality among their children. As Angus Deaton has argued, education is one of the surest paths to achieve the "Great Escape" from the threat of illness and premature mortality (158).

FUTURE DIRECTIONS

In this chapter, we have highlighted the state of empirical evidence linking schooling and income to health. We see four important future directions for research on SES and health: increasing emphasis on establishing causality; improved correspondence between SES measures as operationalized

in research and SES measures as they can be modified with actual interventions and policy changes; better integration of research findings into general theories of how SES affects health; and more formal analyses of the cost-benefit trade-offs of proposed interventions, incorporating both financial and population health benefits.

It is challenging to conduct experiments in the areas of education and income, not least because of ethical concerns, costs, and the likely lag times involved between exposure and changes in health outcomes. Nonetheless, it is critical that we improve the evidence for causality by leveraging quasi-experiments—and, increasingly, real experiments such as charter school lotteries. Experimental and quasi-experimental (IV) studies have often been hamstrung by limited statistical power due to small sample sizes and relatively small effects of policy changes. Null results from studies with limited statistical power are often misinterpreted as evidence of "no effect," even when the imprecise effect estimates are consistent with either large benefits or large harms. Even studies with huge sample sizes may be underpowered to detect plausible effects of an intervention if the intervention under consideration, for example, a CSL change, induced only a small change in social circumstances or was only relevant for a small fraction of the population. Technical innovations linking large surveillance data sets to experimentally or quasi-experimentally assigned exposures may help with addressing this challenge. Joint assessments of multiple outcomes, for example, lumping multiple "noisy" indicators of health together, has proven helpful in evaluations of the early education interventions, with the caveat that not all health outcomes will necessarily move in the same direction. Meta-analyses of results may also prove valuable, although to date there are rarely enough studies of similar exposures to justify meta-analyses providing pooled effect estimates.

Greater reliance on experiments and quasi-experiments will naturally lead to better correspondence between the SES measures in research and realistic interventions; however, this correspondence should also be aggressively pursued in conventional observational research. Observational cohort studies are the bread and butter of epidemiology and, because they are (relatively) inexpensive and flexible, they are likely to remain important sources of evidence in the foreseeable future. However, as we move forward with observational studies, we need to constantly evaluate against experimental evidence to understand sources of bias. For example, charter school evaluations can compare lottery-based results to analyses using matching or regression adjustment approaches to estimate effects (159). There is a tradition of such evaluations in epidemiology, comparing RCT results to observational studies (160, 161), but this should be a priority in social epidemiology. Perhaps even more importantly, experimental results indicate the inadequacy of typical SES measures used in observational studies. Education is not reducible to "years of completed schooling." The most important educational intervention evaluated to date—preschool attendance—would not be included in the measure of education used in nearly all epidemiologic studies of education and health. Some scholars complain that the search for causal evidence has led the field into a quagmire of technical fixes (fancy econometric methods) and experiments that yield little insight into the mechanisms by which SES matters for health (see for example, Angus Deaton's withering critique of IV estimation and field experiments) (162). Nevertheless, there is a middle ground where experimental evidence can provide insights that are useful for policy translation. Theoretical understanding of the mechanisms linking social conditions and health outcomes allows us to generalize from specific contexts of experimental interventions to new

populations and new variations on the experimental treatments. Theoretical considerations should inform both the design and interpretation of experiments, and experimental findings should be fed back to refine our theoretical understanding of how SES affects health. After all, even if we hold strong prior beliefs about causation in general, there is now compelling evidence that the consequences of SES improvements depend on how the new resources are delivered (e.g., increased wages versus lottery winnings), who receives them, and the social context in which the intervention occurs. Numerous questions remain to be answered. The ultimate goal of social epidemiology is to generate actionable information for policy translation. For example, if we agree that the relationship between schooling and health is at least partially causal, then where should society invest—subsidizing preschool programs, or encouraging high school graduation, or expanding access to community colleges? What is the best strategy to transfer income to the poor, so that short-term unintended consequences can be minimized and parents can be incentivized to invest in the future of their children, thereby breaking the intergenerational transmission of poverty? How does the impact of educational interventions change as individuals grow older, and do all individuals benefit similarly? What aspect of education is most important—knowledge, flexible cognitive skills, or simply the social networks obtained through education?

Alongside better understanding of the causal mechanisms, more comprehensive analyses of the population health impact and the net financial consequences of specific strategies to improve health using SES-based strategies would help guide translation efforts. In Chapter 12 we discuss a model of evidence-based public health, in which the final step is evaluating the magnitude of effects of proposed interventions. A major challenge of good programs is taking them to scale, and to understand whether that is feasible requires a simultaneous consideration of possible population health impacts and total costs. Good health is an important outcome regardless of the financial consequences. It may therefore seem irrelevant or even immoral to evaluate the cost of health-promoting interventions. However, if we believe that extra money would improve people's health, the cost of potential health interventions must be considered against the benchmark of simply giving the intended beneficiaries the money. Evidence of favorable cost-benefit profiles can help motivate interventions that improve population health and evaluate potential trade-offs against other desirable social investments.

CONCLUSION

Converging evidence from diverse study designs suggests that socioeconomic conditions are powerful determinants of health. Recent findings, however, suggest a much more complicated picture in which some interventions that seem appealing have trivial or even adverse consequences. A huge evidence gap remains in order to translate the broad findings from studies of the SES-health association into effective strategies to eliminate social inequalities in health. The next generation of researchers will hopefully address these gaps by further evaluating the impact of policies and interventions to understand who benefits, what resources are most helpful, and when and how these resources should be delivered.

REFERENCES

1. Adler NE, Rehkopf DH. US disparities in health: descriptions, causes, and mechanisms. Annu Rev Public Health. 2008;29:235–52.
2. Adler NE, Stewart J. Health disparities across the lifespan: meaning, methods, and mechanisms. Ann N Y Acad Sci. 2010;1186(1):5–23.
3. Krieger N. Epidemiology and the people's health: theory and context. Oxford: Oxford University Press; 2011.
4. Dickens C. Hard times. New York: T. L. McElrath & Co.; 1854.
5. Wharton E. The house of mirth. London, U.K.: The Macmillan Company; 1905.
6. Boo K. Behind the beautiful forevers. 1st ed. New York: Random House; 2012.
7. Davey Smith G, Neaton JD, Wentworth D, Stamler R, Stamler J. Socioeconomic differentials in mortality risk among men screened for the Multiple Risk Factor Intervention Trial: I. White men. Am J Public Health. 1996;86(4):486–96.
8. Link BG, Phelan J. Social conditions as fundamental causes of disease. J Health Soc Behav. 1995;Spec:80–94.
9. Berkman LF, Kawachi I, editors. Social Epidemiology. 1st ed. New York: Oxford University Press, Inc.; 2000.
10. Case A, Paxson C. Parental behavior and child health. Health Aff (Millwood). 2002;21(2):164–78.
11. Farrell P, Fuchs VR. Schooling and health: the cigarette connection. J Health Econ. 1982;1(3):217–30.
12. Gilman SE, Martin LT, Abrams DB, Kawachi I, Kubzansky L, Loucks EB, et al. Educational attainment and cigarette smoking: a causal association? Int J Epidemiol. 2008;37(3):615–24.
13. Chang VW, Lauderdale DS. Fundamental cause theory, technological innovation, and health disparities: the case of cholesterol in the era of statins. J Health Soc Behav. 2009;50(3):245–60.
14. Phelan JC, Link BG, Tehranifar P. Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. J Health Soc Behav. 2010;51(1 Suppl):S28–S40.
15. Fortson JG. The gradient in sub-Saharan Africa: socioeconomic status and HIV/AIDS. Demography. 2008;45(2):303–22.
16. Phelan JC, Link BG, Diez-Roux A, Kawachi I, Levin B. "Fundamental causes" of social inequalities in mortality: a test of the theory. J Health Soc Behav. 2004;45(3):265–85.
17. Smith JP. The impact of socioeconomic status on health over the life-course. J Hum Resour. 2007;42(4):739–64.
18. Galama T, van Kippersluis H. A Theory of Socioeconomic Disparities in Health Over the Life Cycle. RAND Corporation Publications Department, Working Papers: 773, 2010.
19. Grusky DB. The contours of social stratification. In: Grusky DB, editor. Social stratification in sociological perspective. Boulder, CO: Westview Press; 1994. p. 3–35.
20. Krieger N, Rehkopf DH, Chen JT, Waterman PD, Marcelli E, Kennedy M. The fall and rise of US inequities in premature mortality: 1960–2002. PLoS Med. 2008;5(2):e46.
21. Mackenbach JP, Stirbu I, Roskam AJ, Schaap MM, Menvielle G, Leinsalu M, et al. Socioeconomic inequalities in health in 22 European countries. N Engl J Med. 2008;358(23):2468–81.
22. Lopez-Arana S, Burdorf A, Avendano M. Trends in overweight by educational level in 33 low- and middle-income countries: the role of parity, age at first birth and breastfeeding. Obes Rev. 2013;14(10):806–17.
23. Avendano M, Kunst AE, van Lenthe F, Bos V, Costa G, Valkonen T, et al. Trends in socioeconomic disparities in stroke mortality in six European countries between 1981–1985 and 1991–1995. Am J Epidemiol. 2005;161(1):52–61.
24. Avendano M, Kunst AE, Huisman M, Lenthe FV, Bopp M, Regidor E, et al. Socioeconomic status and ischaemic heart disease mortality in 10 western European populations during the 1990s. Heart. 2006;92(4):461–7.

25. Meara ER, Richards S, Cutler DM. The gap gets bigger: changes in mortality and life expectancy, by education, 1981–2000. Health Aff (Millwood). 2008;27(2):350–60.

26. Simmel G. Soziologie: Untersuchungen über die formen der vergesellschaftung. Leipzig: Verlag von Duncker & Humblot; 1908.

27. Sørensen AB. The basic concepts of stratification research: class, status, and power. In: Grusky DB, editor. Social stratification in sociological perspective. Boulder, CO: Westview Press; 1994. p. 229–41.

28. Lynch J, Kaplan G. Socioeconomic Position. In: Berkman LF, Kawachi I, editors. Social epidemiology. 1st ed. New York: Oxford University Press, Inc.; 2000. p. 13–35.

29. Goldthorpe JH, Jackson M. Intergenerational class mobility in contemporary Britain: political concerns and empirical findings. Br J Sociol. 2007;58(4):525–46.

30. US Department of Health and Human Services. Preventing Tobacco Use Among Youth and Young Adults: A Report of the Surgeon General Atlanta: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2012.

31. Kuh D, Ben-Shlomo Y, editors. A lifecourse approach to chronic disease epidemiology: tracing the origins of ill-health from early to adult life. Oxford: Oxford University Press; 1997.

32. Mishra G, Nitsch D, Black S, De Stavola B, Kuh D, Hardy R. A structured approach to modelling the effects of binary exposure variables over the life course. Int J Epidemiol. 2009;38(2):528–37.

33. Nandi A, Glymour M, VanderWeele T. Using marginal structural models to estimate the direct effect of adverse childhood social conditions on onset of heart disease, diabetes, and stroke. Epidemiology. 2012;23(2):223–32.

34. Galobardes B, Smith GD, Lynch JW. Systematic review of the influence of childhood socioeconomic circumstances on risk for cardiovascular disease in adulthood. Ann Epidemiol. 2006;16(2):91–104.

35. Pollitt RA, Rose KM, Kaufman JS. Evaluating the evidence for models of life course socioeconomic factors and cardiovascular outcomes: a systematic review. BMC Public Health. 2005;5:7.

36. Marin TJ, Chen E, Miller GE. What do trajectories of childhood socioeconomic status tell us about markers of cardiovascular health in adolescence? Psychosom Med. 2008;70(2):152–9.

37. Brody GH, Yu T, Chen E, Miller GE, Kogan SM, Beach SR. Is resilience only skin deep? Rural African Americans' socioeconomic status-related risk and competence in preadolescence and psychological adjustment and allostatic load at age 19. Psychol Sci. 2013;24(7):1285–93.

38. James SA. John Henryism and the health of African-Americans. Cult Med Psychiatry. 1994;18(2):163–82.

39. Pearl J. Causality. Cambridge, UK: Cambridge University Press; 2000.

40. National Center for Health Statistics. Health, United States, 2011: With special feature on socioeconomic status and health. Hyattsville, MD: 2012.

41. Kawachi I, Adler NE, Dow WH. Money, schooling, and health: mechanisms and causal evidence. Ann N Y Acad Sci. 2010;1186(1):56–68.

42. Ozer EJ, Fernald LC, Weber A, Flynn EP, VanderWeele TJ. Does alleviating poverty affect mothers' depressive symptoms? A quasi-experimental investigation of Mexico's Oportunidades programme. Int J Epidemiol. 2011;40(6):1565–76.

43. Barrett CB, Carter MR. The power and pitfalls of experiments in development economics: some non-random reflections. Applied Economic Perspectives and Policy. 2010;32(4):515–48.

44. Black D, Working Group on Inequalities in Health. Inequalities in health: The Black report. Department of Health and Social Security; 1980.

45. Case A, Fertig A, Paxson C. The lasting impact of childhood health and circumstance. J Health Econ. 2005;24(2):365–89.

46. Glass CM, Haas SA, Reither EN. The skinny on success: body mass, gender and occupational standing across the life course. Soc Forces. 2010;88(4):1777–806.

47. Ogden C, Lamb M, Caroll M, Flegal K. Obesity and socioeconomic status in adults: United States, 2005–2008. Hyattsville, MD: National Center for Health Statistics, 2010.
48. Conley D, Glauber R. Gender, body mass, and socioeconomic status: new evidence from the PSID. Adv Health Econ Health Serv Res. 2007;17:253–75.
49. Roehling MV. Weight-based discrimination in employment: psychological and legal aspects. Pers Psychol. 1999;52(4):969–1016.
50. Strully KW, Rehkopf DH, Xuan Z. Effects of prenatal poverty on infant health: state earned income tax credits and birth weight. Am Sociol Rev. 2010;75(4):534–62.
51. Fujiwara T, Kawachi I. Is education causally related to better health? A twin fixed-effect study in the USA. Int J Epidemiol. 2009;38(5):1310–22.
52. Angrist J, Pischke J. Mostly harmless econometrics: an empiricist's companion. Princeton, NJ: Princeton University Press; 2009.
53. Angrist JD, Krueger AB. Does compulsory school attendance affect schooling and earnings? Quarterly Journal of Economics. 1991;106(4):979–1014.
54. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology. 1999;10(1): 37–48.
55. Glymour MM. Natural experiments and instrumental variables analyses in social epidemiology. In: Oakes JM, Kaufman JS, editors. Methods in social epidemiology. San Francisco: Jossey-Bass; 2006.
56. Angrist JD, Imbens GW. 2-stage least-squares estimation of average causal effects in models with variable treatment intensity. J Am Stat Assoc. 1995;90(430):431–42.
57. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. J Am Stat Assoc. 1996;91(434):444–55.
58. Kitigawa EM, Hauser PM. Differential mortality in the United States: a study in socioeconomic epidemiology. Cambridge, MA: Harvard University Press; 1973.
59. Elo IT, Preston SH. Educational differentials in mortality: United States, 1979–1985. Soc Sci Med. 1996;42(1):47–57.
60. Huisman M, Kunst AE, Andersen O, Bopp M, Borgan JK, Borrell C, et al. Socioeconomic inequalities in mortality among elderly people in 11 European populations. J Epidemiol Community Health. 2004;58(6): 468–75.
61. Meara ER, Richards S, Cutler DM. The gap gets bigger: changes in mortality and life expectancy, by education, 1981–2000. Health Aff (Millwood). 2008;27(2):350–60.
62. Huisman M, Kunst AE, Bopp M, Borgan JK, Borrell C, Costa G, et al. Educational inequalities in cause-specific mortality in middle-aged and older men and women in eight western European populations. Lancet. 2005;365(9458):493–500.
63. Manly JJ, Touradji P, Tang MX, Stern Y. Literacy and memory decline among ethnically diverse elders. J Clin Exp Neuropsychol. 2003;25(5):680–90.
64. Glymour MM, Manly JJ. Lifecourse social conditions and racial and ethnic patterns of cognitive aging. Neuropsychol Rev. 2008;18(3):223–54.
65. Liu SY, Linkletter CD, Loucks EB, Glymour MM, Buka SL. Decreased births among black female adolescents following school desegregation. Soc Sci Med. 2012;74(7):982–8.
66. Johnson RC. Long-run impacts of school desegregation and school quality on adult attainments. National Bureau of Economic Research, 2011.
67. Frisvold D, Golberstein E. The effect of school quality on black-white health differences: evidence from segregated southern schools. Demography. 2013;50(6):1989–2012.
68. Chetty R, Friedman JN, Hilger N, Saez E, Schanzenbach DW, Yagan D. How does your kindergarten classroom affect your earnings? Evidence from Project Star. The Quarterly Journal of Economics. 2011;126(4):1593–660.

69. Liu SY, Chavan NR, Glymour MM. Type of high-school credentials and older age ADL and IADL limitations: is the GED credential equivalent to a diploma? Gerontologist. 2013;53(2):326–33.

70. Caputo RK. The GED as a predictor of mid-life health and economic well-being. Journal of Poverty. 2005;9(4):73–97.

71. Caputo RK. The GED as a signifier of later life health and economic well-being. Race, Gender and Class. 2005;12(2):81–103.

72. Zajacova A. Health in working-aged Americans: Adults with high school equivalency diploma are similar to dropouts, not high school graduates. Am J Public Health. 2012;102(S2):284–90.

73. Zajacova A, Everett BG. The nonequivalent health of high school equivalents. Social Science Quarterly. 2013;95:221–238.

74. O'Donnell K. Adult education participation in 2004–05 (NCES 2006–077). Washington, DC: US Department of Education, National Center for Education Statistics, 2006.

75. Stern C, Munn Z. Cognitive leisure activities and their role in preventing dementia: a systematic review. Int J Evid Based Healthc. 2010;8(1):2–17.

76. Knudsen EI, Heckman JJ, Cameron JL, Shonkoff JP. Economic, neurobiological, and behavioral perspectives on building America's future workforce. Proc Natl Acad Sci U S A. 2006;103(27):10155–62.

77. Card D. Estimating the return to schooling: progress on some persistent econometric problems. Econometrica. 2001;69(5):1127–60.

78. Kramer AF, Bherer L, Colcombe SJ, Dong W, Greenough WT. Environmental influences on cognitive and brain plasticity during aging. J Gerontol A Biol Sci Med Sci. 2004;59(9):940–57.

79. Snyder HN, Sickmund M. Juvenile offenders and victims: 2006 national report. Office of Juvenile Justice and Delinquency Prevention, 2006.

80. Belfield CR, Nores M, Barnett S, Schweinhart L. The High/Scope Perry Preschool Program cost–benefit analysis using data from the age-40 followup. J Hum Resour. 2006;41(1):162–90.

81. Muennig P, Schweinhart L, Montie J, Neidell M. Effects of a prekindergarten educational intervention on adult health: 37-year follow-up results of a randomized controlled trial. Am J Public Health. 2009;99(8):1431–7.

82. Campbell FA, Ramey CT, Pungello EP, Sparling JJ, Miller-Johnson S. Early Childhood Education: Young Adult Outcomes from the Abecedarian Project. Appl Dev Sci. 2002;6:42–57.

83. McLaughlin AE, Campbell FA, Pungello EP, Skinner M. Depressive symptoms in young adults: the influences of the early home environment and early educational child care. Child Dev. 2007;78(3):746–56.

84. Muennig P, Robertson D, Johnson G, Campbell F, Pungello EP, Neidell M. The effect of an early education program on adult health: the Carolina Abecedarian Project randomized controlled trial. Am J Public Health. 2011;101(3):512–6.

85. Muennig P, Johnson G, Wilde ET. The effect of small class sizes on mortality through age 29 years: evidence from a multicenter randomized controlled trial. Am J Epidemiol. 2011;173(12):1468–74.

86. Wilde ET, Finn J, Johnson G, Muennig P. The effect of class size in grades K-3 on adult earnings, employment, and disability status: evidence from a multi-center randomized controlled trial. J Health Care Poor Underserved. 2011;22(4):1424–35.

87. Barnett WS. Effectiveness of early educational intervention. Science. 2011;333(6045):975–8.

88. Puma M, Bell S, Cook R, Heid C, Shapiro G, Broene P, et al. Head Start impact study: final report. Administration for Children and Families, 2010. http://eclkc.ohs.acf.hhs.gov/hslc/mr/factsheets/docs/hs-program-fact-sheet-2012.pdf.

89. Gibbs C, Ludwig J, Miller DL. Does Head Start do any lasting good? NBER Working Paper 17452. 2011;NBER Working Paper Series.

90. Garces E, Thomas D, Currie J. Longer-term effects of Head Start. Am Econ Rev. 2002;92(4):999–1012.

91. Deming D. Early childhood intervention and life-cycle skill development: evidence from Head Start. American Economic Journal: Applied Economics. 2009:111–34.
92. Anderson KH, Foster JE, Frisvold DE. Investing in health: the long-term impact of Head Start on smoking. Econ Inq. 2010;48(3):587–602.
93. Ludwig J, Miller DL. Does Head Start improve children's life chances? Evidence from a regression discontinuity design. Quarterly Journal of Economics. 2007;122(1):159–208.
94. Barnett WS. Surprising agreement on Head Start: compli/ementing Currie and Besharov. J Policy Anal Manage. 2007;26(3):685–6.
95. Lleras-Muney A. Were compulsory attendance and child labor laws effective? An analysis from 1915 to 1939. Journal of Law and Economics. 2002;45(2):401–35.
96. Glymour MM, Kawachi I, Jencks CS, Berkman LF. Does childhood schooling affect old age memory or mental status? Using state schooling laws as natural experiments. J Epidemiol Community Health. 2008;62(6):532–7.
97. Banks J, Mazzonna F. The effect of education on old age cognitive abilities: evidence from a regression discontinuity design. Economic Journal. 2012;122:418–48.
98. Schneeweis N, Skirbekk V, Winter-Ebmer R. Does schooling improve cognitive functioning at older ages? Social Science Research Network, 2012.
99. Lleras-Muney A. The relationship between education and adult mortality in the US. Review of Economic Studies. 2005;72(1):189–221.
100. Clark D, Royer H. The effect of education on adult mortality and health: evidence from Britain. Am Econ Rev. 2013;103(6):2087–120.
101. Lager ACJ, Torssander J. Causal effect of education on mortality in a quasi-experiment on 1.2 million Swedes. Proc Natl Acad Sci U S A. 2012;109(22):8461–6.
102. Silles MA. The intergenerational effect of parental education on child health: evidence from the UK. Education Economics. 2013(ahead-of-print):1–15.
103. Albouy V, Lequien L. Does compulsory education lower mortality? J Health Econ. 2009;28(1):155–68.
104. Gathmann C, Jürges H, Reinhold S. Compulsory schooling reforms, education and mortality in twentieth century Europe. CESifo Working Paper Series No. 3755, 2012.
105. Bingley P, Kristensen N. Historical schooling expansions as instruments 2013. Available from: http://www.nhh.no/Admin/Public/DWSDownload.aspx?File=%2FFiles%2FFiler%2Finstitutter%2Fsam%2FConferences%2FNordic+Econometrics+2013%2FBingley-kristensen-1937-reform-20130315.pdf.
106. Snyder T, Dillow S. Digest of education statistics 2011 (NCES 2012–001). Washington, DC: US Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2012.
107. Olshansky SJ, Antonucci T, Berkman L, Binstock RH, Boersch-Supan A, Cacioppo JT, et al. Differences in life expectancy due to race and educational differences are widening, and many may not catch up. Health Aff (Millwood). 2012;31(8):1803–13.
108. Montez JK, Zajacova A. Trends in mortality risk by education level and cause of death among US white women from 1986 to 2006. Am J Public Health. 2013;103(3):473–9.
109. Backlund E, Sorlie PD, Johnson NJ. The shape of the relationship between income and mortality in the United States: evidence from the National Longitudinal Mortality Study. Ann Epidemiol. 1996;6(1):12–20; discussion 1–2.
110. Ecob R, Smith GD. Income and health: what is the nature of the relationship? Soc Sci Med. 1999;48(5):693–705.
111. Mackenbach JP, Martikainen P, Looman CW, Dalstra JA, Kunst AE, Lahelma E. The shape of the relationship between income and self-assessed health: an international study. Int J Epidemiol. 2005;34(2):286–93.
112. Martikainen P, Makela P, Koskinen S, Valkonen T. Income differences in mortality: a register-based follow-up study of three million men and women. Int J Epidemiol. 2001;30(6):1397–405.

113. Rahkonen O, Arber S, Lahelma E, Martikainen P, Silventoinen K. Understanding income inequalities in health among men and women in Britain and Finland. Int J Health Serv. 2000;30(1):27–47.

114. Galama T, van Kippersluis H. Health inequalities through the lens of health capital theory: issues, solutions, and future directions. Santa Monica, CA: RAND Corporation, 2013.

115. Anderson LM, Shinn C, Fullilove MT, Scrimshaw SC, Fielding JE, Normand J, et al. The effectiveness of early childhood development programs: a systematic review. Am J Prev Med. 2003;24(3 Suppl):32–46.

116. Doyle O, Harmon CP, Heckman JJ, Tremblay RE. Investing in early human development: timing and economic efficiency. Econ Hum Biol. 2009;7(1):1–6.

117. Feinstein L. Inequality in the early cognitive development of British children in the 1970 cohort. Economica. 2003;70(277):73–97.

118. Halfon N, Hochstein M. Life course health development: an integrated framework for developing health, policy, and research. Milbank Q. 2002;80(3):433–79, iii.

119. Halfon N, Larson K, Lu M, Tullis E, Russ S. Lifecourse health development: past, present and future. Matern Child Health J. 2013.

120. Duncan GJ, Morris PA, Rodrigues C. Does money really matter? Estimating impacts of family income on young children's achievement with data from random-assignment experiments. Dev Psychol. 2011;47(5):1263–79.

121. Desai S, Chase-Lansdale PL, Michael RT. Mother or market? Effects of maternal employment on the intellectual ability of 4-year-old children. Demography. 1989;26(4):545–61.

122. Becker G. A theory of the allocation of time. Economic Journal. 1906;75(299):493–517.

123. Gennetian LA, Miller C. Children and welfare reform: a view from an experimental welfare program in Minnesota. Child Dev. 2002;73(2):601–20.

124. Cooper K, Steward K. Does money affect children's outcomes? A systematic review. London: London School of Economics & Joseph Rowntree Foundation, 2013. http://www.jrf.org.uk/sites/files/jrf/money-children-outcomes-full.pdf.

125. Clark-Kauffman E, Duncan GJ, Morris P. How welfare policies affect child and adolescent achievement. Am Econ Rev. 2003;93(2):299–303.

126. Meyer BD, Rosenbaum DT. Welfare, the earned income tax credit, and the labor supply of single mothers. Quarterly Journal of Economics. 2001;116(3):1063–114.

127. Dahl GB, Lochner L. The impact of family income on child achievement: evidence from the earned income tax credit. Am Econ Rev. 2012;102(5):1927–56.

128. Cowan B, Tefft N. Education, maternal smoking, and the earned income tax credit. BE Journal of Economic Analysis and Policy. 2012;13(1):1.

129. Averett S, Wang Y. The effects of earned income tax credit payment expansion on maternal smoking. Health Econ. 2013;22(11):1344–59.

130. Larrimore J. Does a higher income have positive health effects? Using the earned income tax credit to explore the income-health gradient. Milbank Q. 2011;89(4):694–727.

131. Schmeiser MD. Expanding wallets and waistlines: the impact of family income on the BMI of women and men eligible for the earned income tax credit. Health Econ. 2009;18(11):1277–94.

132. Pega F, Carter K, Kawachi I, Davis P, Gunasekara FI, Lundberg O, et al. The impact of in-work tax credit for families on self-rated health in adults: a cohort study of 6900 New Zealanders. J Epidemiol Community Health. 2013;67(8):682–8.

133. Fiszbein A, Schady N, Ferreira F, Grosh M, Keleher N, Olinto P, et al. Conditional cash transfers: reducing present and future poverty. Washington: The International Bank for Reconstruction and Development/The World Bank, 2009.

134. Rasella D, Aquino R, Santos CA, Paes-Sousa R, Barreto ML. Effect of a conditional cash transfer programme on childhood mortality: a nationwide analysis of Brazilian municipalities. Lancet. 2013;382(9886):57–64.

135. Shei A. Brazil's conditional cash transfer program associated with declines in infant mortality rates. Health Aff (Millwood). 2013;32(7):1274–81.

136. Forde I, Chandola T, Garcia S, Marmot MG, Attanasio O. The impact of cash transfers to poor women in Colombia on BMI and obesity: prospective cohort study. Int J Obes (Lond). 2012;36(9):1209–14.

137. Andalon M. Oportunidades to reduce overweight and obesity in Mexico? Health Econ. 2011;20(1 Suppl):1–18.

138. Leroy JL, Gadsden P, Rodriguez-Ramirez S, de Cossio TG. Cash and in-kind transfers in poor rural communities in Mexico increase household fruit, vegetable, and micronutrient consumption but also lead to excess energy consumption. J Nutr. 2010;140(3):612–7.

139. Lindahl M. Estimating the effect of income on health using lottery prizes as exogenous sources of variation in income. J Hum Resour. 2005;40(1):144–68.

140. Apouey B, Clark A. Winning big but feeling no better? The effect of lottery prizes on physical and mental health. Working paper No. 2009–09, Paris School of Economics, 2009.

141. Meer J, Miller DL, Rosen HS. Exploring the health-wealth nexus. J Health Econ. 2003;22(5):713–30.

142. Michaud PC, van Soest A. Health and wealth of elderly couples: causality tests using dynamic panel data models. J Health Econ. 2008;27(5):1312–25.

143. Kim B, Ruhm CJ. Inheritances, health and death. Health Econ. 2012;21(2):127–44.

144. Kuhn P, Kooreman P, Soetevent A, Kapteyn A. The effects of lottery prizes on winners and their neighbors: evidence from the Dutch postcode lottery. Am Econ Rev. 2011;101(5):2226–47.

145. Costello EJ, Compton SN, Keeler G, Angold A. Relationships between poverty and psychopathology: a natural experiment. JAMA. 2003;290(15):2023–9.

146. Costello EJ, Erkanli A, Copeland W, Angold A. Association of family income supplements in adolescence with development of psychiatric and substance use disorders in adulthood among an American Indian population. JAMA. 2010;303(19):1954–60.

147. Akee R, Simeonova E, Copeland W, Angold A, Costello JE. Young adult obesity and household income: effects of unconditional cash transfers. American Economic Journal: Applied Economics. 2013;5(2):1–28.

148. Wolfe B, Jakubowski J, Haveman R, Courey M. The income and health effects of tribal casino gaming on American Indians. Demography. 2012;49(2):499–524.

149. Bruckner TA, Brown RA, Margerison-Zilko C. Positive income shocks and accidental deaths among Cherokee Indians: a natural experiment. Int J Epidemiol. 2011;40(4):1083–90.

150. Kaleta D, Makowiec-Dabrowska T, Dziankowska-Zaborszczyk E, Fronczak A. Predictors of smoking initiation—Results from the Global Adult Tobacco Survey (GATS) in Poland 2009–2010. Ann Agric Environ Med. 2013;20(4):756–66.

151. Novo M, Hammarstrom A, Janlert U. Smoking habits—a question of trend or unemployment? A comparison of young men and women between boom and recession. Public Health. 2000;114(6):460–3.

152. Hammarstrom A, Janlert U. Unemployment—an important predictor for future smoking: a 14-year follow-up study of school leavers. Scand J Public Health. 2003;31(3):229–32.

153. Blakely T, van der Deen FS, Woodward A, Kawachi I, Carter K. Do changes in income, deprivation, labour force status and family status influence smoking behaviour over the short run? Panel study of 15 000 adults. Tob Control. 2013.

154. Xu X. The business cycle and health behaviors. Soc Sci Med. 2012;77:126–36.

155. Cutler DM, Lleras-Muney A. Education and health: insights from international comparisons. 2012.

156. De Walque D. Does education affect HIV status? Evidence from five African countries. World Bank Economic Review. 2009;23(2):209–33.

157. Preston SH, Haines MR. Fatal years: child mortality in late nineteenth-century America. Princeton, NJ: Princeton University Press; 1991.

158. Deaton A. The great escape: health, wealth, and the origins of inequality. Princeton, NJ: Princeton University Press; 2013.

159. Abdulkadiroğlu A, Angrist JD, Dynarski SM, Kane TJ, Pathak PA. Accountability and flexibility in public schools: evidence from Boston's charters and pilots. Quarterly Journal of Economics. 2011;126(2): 699–748.

160. Pocock SJ, Elbourne DR. Randomized trials or observational tribulations?[comment]. N Engl J Med. 2000;342(25):1907–9.

161. Ioannidis JP, Haidich AB, Lau J. Any casualties in the clash of randomised and observational evidence? BMJ. 2001;322(7291):879–80.

162. Deaton A. Instruments, randomization, and learning about development. J Econ Lit. 2010;48(2):424–55.

CHAPTER 3

DISCRIMINATION AND HEALTH INEQUITIES

Nancy Krieger

> Our future survival is predicated upon our ability to relate within equality.
>
> Audre Lorde, 1980

Inequality hurts. Discrimination harms health. These seem like straightforward, even self-evident, statements. They are propositions that epidemiologists can test, just like any other proposition about health that we investigate.

When I wrote the above paragraph, in 1999, in the first published epidemiologic review article on discrimination and health (1), empirical research on discrimination as a determinant of population health was in its infancy.

At that time, I could identify only 20 studies in the public health literature that employed instruments to measure self-reported experiences of discrimination. Of these, 15 focused on racial discrimination (13 on African Americans, two on Hispanics and Mexican Americans), two of which additionally addressed gender discrimination; another solely examined gender discrimination; three investigated discrimination based on sexual orientation; and one concerned discrimination based on disability; all were from the United States; none addressed discrimination based on age.

Since then, research has burgeoned. As I discuss below, the number of empirical studies that directly measure exposure to discrimination to analyze its links to health easily exceeds 500, as tallied up in review articles, with these studies increasingly global in scope and focused on major types of discrimination (Table 3.1) variously involving race/ethnicity, Indigenous status, immigrant status, gender, sexuality, disability, and age, separately and in combination. And yet, as I will also document, even as the number of investigations has dramatically expanded, the scope remains narrow. The overwhelming emphasis is on interpersonal discrimination, referring to encounters between individuals in which one person acts in an adversely discriminatory way towards another person, with this type of discrimination primarily conceptualized as a psychosocial stressor, and on the

TABLE 3.1: Basic taxonomy of prevalent types of discrimination, United States: by: type, constituent dominant and subordinated social groups, justifying ideology, material and social basis, and measurable aspects

Type of Discrimination	Constituent Social Groups		Justifying Ideology	Material and Social Basis	Measurable Aspects (Common to Each Type of Discrimination)
	Dominant	Subordinated			
Racial/ethnic	White, Euro-American	People of color[a]: Black or African American; American Indian or Alaska Native; Asian; Native Hawaiian or Other Pacific Islander; Hispanic or Latino	Racism	Conquest, slavery, skin color, property	Expression of Discrimination Form: legal or illegal; institutional, structural, interpersonal; direct or indirect; overt or covert Agency: perpetrated by state or by nonstate actors (institutional or individuals) Expression: from verbal to violent; mental, physical, or sexual Domain: e.g., at home; within family; at school; getting a job; at work; getting housing; getting credit or loans; getting medical care, purchasing other goods and services; by the media; from the police or in the courts; by other public agencies or social services; on the street or in a public setting Level: individual, institutional, residential neighborhood, political jurisdiction, regional economy Cumulative Exposure to Discrimination Timing: conception; infancy; childhood; adolescence; adulthood Intensity Frequency (acute; chronic) Duration
Anti-immigrant (see also racism for immigrants of color)	Native-born citizen	Foreign-born immigrants, both legal and undocumented	Nativism	Labor market, language	

					Responses to Discrimination (Protective and Harmful) Protective Active resistance by individuals and communities (e.g., involving organizing, lawsuits, social networks, social support) Creating safe spaces for self-affirmation (e.g., social, cultural, sexual) Harmful Internalized oppression and denial Use of psychoactive substances (legal and illegal)
Gender[b]	Men and boys	Women and girls	Sexism	Property, gender roles, religion	
Anti-gay, lesbian, bisexual, transgender (LGBT)	Heterosexual	Lesbian, gay, bisexual, queer, transgender, transsexual	Heterosexism	Gender roles, religion	
Disability	Able-bodied	Disabled	Ableism	Costs of enabling access	
Age	Nonretired adults	Youth, elderly	Ageism	Family roles, property	
Social class	Business owners, executives, professionals	Working class, poor	Class bias	Property, Education	

[a] Each of these racial/ethnic groups is extremely heterogeneous; the terms listed are the major classifications employed since 1997 by the US government, including in the census (Office of Management and Budget (175)). Examples (far from exhaustive) of subgroups include: Black: African American, Afro-Caribbean, and Black African; Latino/a and Hispanic: Chicano, Mexican American, Cuban, Puerto Rican, Central and South American; Native Hawaiian and Pacific Islander: Native Hawaiian, Samoan, Guamanian; Asian: Chinese, Japanese, Filipino, Korean, Laotian, Hmong, Samoan; American Indian and Alaska Native: 565 federally recognized American Indian tribes and additional state-recognized tribes and tribes not recognized by either the US federal or state governments, and also Aleuts, and Eskimos.
[b] Also called "sex discrimination."

biological consequences of exposure to toxic stress. By contrast, empirical studies on the health impacts of structural discrimination, referring to discrimination enacted by institutions (e.g., laws or rules that impose adverse discrimination, by design, such as legalized racial discrimination, or in effect, such as the racialized impact of the New York Police Department's "stop-and-frisk" policy, under legal contest at the time of preparing this review [2]), remain scant, a gap consonant with the limited epidemiologic research on political systems and population health (3, 4).

The individual-level approach to analyzing discrimination and health, moreover, coexists with the still dominant biomedical orientation, also focused on the individual level, which typically ignores social determinants of health and emphasizes genetic causes of disease within individuals to explain group rates of disease (4, 5). Exemplifying the dominant orientation (6), not only did the National Institute of Health's 2008–2009 biennial report to Congress (7) allocate only 46 of its 732 pages to "Minority Health and Health Disparities" but also within these 46 pages the clear emphasis was on assumed genetic explanations of racial/ethnic differences in health status. Thus, the terms "genome," "genomic," "genetic," and "gene" appear 87 times, whereas "social determinants of health" and "discrimination" each occur once, "socioeconomic" 7 times, "poverty" twice, and "racism" not at all (7, 8).

Of course, knowledge of biological mechanisms is vital for vetting causal claims. Equally essential, however, is research on the societal mechanisms that generate discrimination and the myriad pathways, material as well as psychological, by which it becomes embodied and biologically expressed in individuals' health status and population patterns of health inequities (4, 8). Stated simply, all biological phenomena—including health and disease—involve gene expression; what stands out is the neglect of the societal and ecological context driving this expression. At issue are not only day-to-day living and working conditions but also the exercise of civil, political, economic, social, and cultural rights, that is, human rights (9–11).

The question at hand, after all, is how individual and population health are affected by the economic consequences of discrimination and the accumulated insults arising from everyday and at times violent experiences of being treated as a second-class citizen. Focus solely on experiences that people can self-report, or on what can be gleaned from experimental studies, and neglect analysis of discriminatory exposures that can only be measured at the population level, and the full picture of discrimination's toll recedes from view (8, 12). The totality of evidence of embodied harm, from institutionally to individually imposed, whether or not consciously named as discrimination by the targets of the discriminatory actions, is crucial—both for analyzing etiology and for guiding action to rectify and prevent health inequities.

Accordingly, to help advance the state of the field, this chapter will first briefly review definitions of discrimination, offer illustrations of their patterns within the United States, and discuss theoretical insights useful for conceptualizing how discrimination can become embodied and produce health inequities, including via distortion of scientific knowledge. It then will concisely summarize extant evidence—both robust and inconsistent—linking discrimination and health, after which it will focus on several key methodological controversies and challenges. The examples discussed, while often drawing on US data (which still constitutes the bulk of work), raise conceptual, methodologic, and substantive issues germane for any type of discrimination in any country context.

But first: one critical caveat. The purpose of studying the health consequences of discrimination is not to prove that oppression is "bad" because it harms health. Unjustly denying people

fair treatment, abrogating human rights, and constraining possibilities for living fully expressed, dignified, and loving lives is, by definition, wrong (9, 10, 13, 14)—*regardless* of effects on health. Rather, the rationale for studying discrimination and health, like that for studying any societal determinant of health, is to render an accounting of who and what drives population patterns of health and health inequities (15) and to generate knowledge useful for guiding policies and actions to prevent and rectify harm and advance health equity.

DISCRIMINATION: DEFINITIONS AND PATTERNS

DEFINITIONS OF DISCRIMINATION

According to the *Oxford English Dictionary* (16), the word "discriminate" derives from the Latin term *discriminare*, which means "to divide, separate, distinguish." From this standpoint, "discrimination" simply means "a distinction (made with the mind, or in action)." Yet, when people are involved, as both agents and objects of discrimination, discrimination takes on a new meaning: "to discriminate against" is "to make an adverse distinction with regard to; to distinguish unfavorably from others." In other words, when people belonging to one societal group exclude and discriminate against people outside of their group, more than simple distinctions are at issue. Instead, those who discriminate restrict, by judgment and action, the lives of those against whom they discriminate.

The invidious meanings of adverse discrimination become readily apparent in the legal realm, where people have created and enforce laws both to uphold and to challenge discrimination. Legally, discrimination can be of two forms. One is de jure, meaning mandated by law; the other is de facto, without legal basis but sanctioned by custom or practice. Examples of de jure discrimination in the United States include Jim Crow laws, now overturned, that denied African Americans access to facilities and services used by white Americans (17–19) and entrenched laws, increasingly contested, prohibiting gay and lesbian marriage (20–22). By contrast, differential and inadequate treatment, for example, by race/ethnicity or gender, of persons otherwise medically warranting the same care constitutes a form of de facto discrimination (23, 24).

Whether de jure or de facto, discrimination can be perpetrated by a diverse array of actors. These include the state and its institutions (ranging from law courts to public schools), nonstate entities (e.g., private sector employers, private schools, religious organizations), and individuals. From a legal or human rights perspective, however, it is the state that possesses critical agency and establishes the context—whether permissive or prohibitive—for discriminatory acts: It can enforce, enable, or condone discrimination, or, alternatively, it can outlaw discrimination and seek to redress its effects (Table 3.2) (9, 10, 14). A powerful example of the latter is the post-apartheid South African constitution (25). This document mandates, in the most inclusive language of any national constitution in the world, that "The state may not unfairly discriminate directly or indirectly against anyone on one or more grounds, including race, gender, sex, pregnancy, marital status, ethnic or social origin, colour, sexual orientation, age, disability, religion, conscience, belief, culture, language and birth"; discrimination by individuals on these terms is likewise prohibited.

TABLE 3.2: Selected US laws and international human rights instruments prohibiting discrimination

US Laws	International Human Rights Instruments
US Constitution	Universal Declaration of Human Rights (1948)
13th Amendment (banned slavery) (1865)	Discrimination (Employment and Occupation Convention) (1958)
14th Amendment (guaranteed due process to all citizens, excepting American Indians) (1866)	Convention Against Discrimination (in Education) (1960) International Convention on the Elimination of All Forms of Racial Discrimination (1965)
15th Amendment (banned voting discrimination based on "race, color, or previous condition of servitude") (1870)	International Covenant on Civil and Political Rights (1966)
19th Amendment (banned voting discrimination "on account of sex") (1920)	International Covenant on Economic, Social, and Cultural Rights (1966) Declaration on the Elimination of Discrimination Against Women (1967)
Civil Rights Act (1875) (declared unconstitutional by US Supreme Court in 1883)	Declaration on Race and Racial Prejudice (1978)
Civil Rights Act (1964)	Convention on the Elimination of All Forms of Discrimination Against Women (1979)
Voting Rights Act (1965)	Convention on the Rights of the Child (1989)
Fair Housing Act (1968)	
Equal Opportunity Act (1975)	
Americans with Disability Act (1990)	
Genetic Information Nondiscrimination Act (2008)	
Lilly Ledbetter Fair Pay Act (2009)	
Matthew Shepard and James Byrd, Jr. Hate Crimes Prevention Act (2009)	
Fair Sentencing Act (2010)	

Sources: Jaynes and Williams (18: pp. 224–38); Tomasevski (14); ADA (176); GINA (177); US Government (178).

Even so, as attested to by South Africa's growing economic inequality and persistent racial/ethnic economic and health inequities (26), the legal abolition of contemporary discrimination, however essential, is by itself insufficient to eradicate the lifelong health and social consequences, within and across generations, of past discrimination or to change extant distributions of accumulated power and wealth without additional reform (8, 19, 27).

Despite its legal dimensions, however, discrimination is never simply a legal affair. Conceptualized more broadly, it refers to all means of expressing and institutionalizing social relationships of dominance and oppression. At issue are practices of dominant groups to maintain privileges they accrue through subordinating the groups they oppress and the ideologies they use to justify these practices, which typically revolve around notions of innate superiority and inferiority, difference, or deviance. Thus, the *Collins Dictionary of Sociology* defines "discrimination" as "the process by which a member, or members, of a socially defined group is, or are, treated differently (especially unfairly) because of his/her/their membership of that group" (28: p. 169). Extending this definition, the *Concise Oxford Dictionary of Sociology* holds that discrimination involves not only "socially derived beliefs each [group] holds about the other" but also "patterns of dominance and oppression, viewed as expressions of a struggle for power and privilege" (29: pp. 125–6). In other words, random acts of unfair treatment do not constitute discrimination. Instead, discrimination is a socially structured and sanctioned phenomenon, justified by ideology and expressed in interactions among and between individuals and institutions, that maintains privileges for members of dominant groups at the cost of deprivation for others.

Although sharing a common thread of systemic unfair treatment, discrimination nevertheless can vary in form and type, depending on how it is expressed, by whom, and against whom. As summarized in Table 3.1, diverse forms identified by social scientists include: *legal, illegal, overt* (or *blatant*), and *covert* (or *subtle*) discrimination, and also *institutional* (or *organizational*), *structural* (or *systemic*), and *interpersonal* (or *individual*) discrimination (12, 30, 31). Although usage of these terms varies, *institutional discrimination* typically refers to discriminatory policies or practices carried out by state or nonstate institutions; *structural discrimination* refers to the totality of ways in which societies foster discrimination, via mutually reinforcing systems of discrimination (e.g., in housing, education, employment, earnings, benefits, credit, media, health care, criminal justice, etc.) that in turn reinforce discriminatory beliefs, values, and distribution of resources (32); and *interpersonal discrimination* refers to directly perceived discriminatory interactions between individuals—whether in their institutional roles (e.g., employer/employee) or as public or private individuals (e.g., shopkeeper/shopper). In all cases, perpetrators of discrimination act unfairly toward members of socially defined subordinate groups to reinforce relations of dominance and subordination, thereby bolstering privileges conferred to them as members of a dominant group.

PATTERNS OF DISCRIMINATION: US EXAMPLES

A full accounting of discrimination in the United States today is beyond the scope of this chapter. Instead, to provide a reminder of its ubiquity as well as background to considering how it can harm health, I next review, briefly, five notable ways that discrimination can permeate people's lives.

First, as summarized in Table 3.1, many groups experience discrimination in the United States at present. Dominant types of discrimination are based on race/ethnicity, Indigenous status, immigrant status, gender, sexuality, disability, age, and, although not always recognized as such, social class (20, 31, 33–36). Although each type of discrimination has its own justifying ideology, material basis, and legal history (see Table 3.1), all share the common feature of systematic inequitable treatment directed against and adversely affecting individuals in the subordinated group, to the benefit, at the group level if not the individual level, of those who belong to the dominant group.

Second, as explicitly recognized by the South African constitution, and as theoretically propounded in sociological and legal work on "intersectionality" (31, 37, 38), individuals can experience multiple forms of discrimination. For example, whereas white women may be subject, as women, to gender discrimination, women of color—whether black, Latina, Asian or Pacific Islander, or American Indian—may be subject to both gender and racial discrimination. Moreover, this experience of multiple subordination cannot simply be reduced to the "sum" of each type. During the past two decades, a growing body of scholarship on gendered racism, for example, has elucidated how, in a context of overall negative stereotypical portrayals of black Americans as lazy and unintelligent (30, 31), black women—as *black women*—remain stereotyped, as Patricia Collins observed back in 1990 (39: p. 97), as "mammies, matriarchs, welfare recipients and hot mammas," while black men—as *black men*—remain stereotyped as criminals and rapists (31, 39). Understanding discrimination experienced by black women and men thus requires considering the salience of, minimally, both their race/ethnicity and gender; also germane are their socioeconomic position, sexuality, nativity, and age, as is true for members of any societal group.

Third, singly or combined, different types of discrimination can occur in just about every facet of public and private life. The full gamut extends from the grinding daily realities of what Philomena Essed two decades ago influentially termed "everyday" discrimination (30) to less common yet terrifying and life-transforming events, such as being victim of a hate crime (31, 35).

In a typical day experiences with discrimination accordingly can start—depending on type—in the morning, at home; continue with public encounters en route to or while at school or work or when shopping or eating at a restaurant or attending a public event; and extend on through the evening, whether in the news or entertainment or while engaging with family members (30, 31, 33, 35, 36). Other common but not typically daily scenarios for experiencing discrimination include applying for a job, looking for housing, getting a mortgage or a loan, getting health care, or interacting with the police or public agencies or the legal system (12, 31, 33–35).

Fourth, while some experiences of discrimination may be interpersonal and obvious, they are more likely to be institutional and invisible. To know, for example, that you have been discriminated against in your salary, or that you have been denied a mortgage, or an apartment, or been steered away from certain neighborhoods when you are looking for a home, requires knowing how the employer, bank, landlord, or real estate agent treats other individuals (1, 12). Typically, it is only when people file charges of discrimination in court that evidence of such patterns of inequality can be obtained. Other clues can be obtained by examining social patterning of economic inequality, since acts of discrimination—whether institutional or interpersonal, blatant or covert—usually harm economic as well as social well-being (12, 30–39). Table 3.3 illustrates this point for US racial/ethnic discrimination, depicting marked racial/ethnic inequalities in wealth, poverty, education, unemployment, health insurance, incarceration, and political parity in representation.

Fifth and finally, attesting to some of the animosity that feeds and justifies discrimination are, to give but one example, a long lineage of US. racial attitudes (18, 40). Despite declines in racial prejudice over time, reported levels remain high, even taking into account that: *(1)* people underreport negative social attitudes, and *(2)* dominant groups typically deny discrimination exists (12, 40–42), especially if it is no longer legal (see, for example [43–45]). Indeed, as Jackman has long argued (46), paternalism combined with *(1)* friendly feelings toward individual members of subordinate groups and *(2)* denial of any responsibility for institutional discrimination is as much a hallmark of contemporary discrimination as is outright conflict and

TABLE 3.3: Analyzing US racial/ethnic discrimination and health inequities in context: relevant socioeconomic, sociopolitical, and embodied facts

Racial/Ethnic Inequity Indicators	Total US Population	White Non-Hispanic[a]	Asian[b]	Hispanic	Black Non-Hispanic[c]	American Indian/ Alaska Native
Wealth (2011): net worth ($)	68,828	110,500	89,399	7,683	6,314	…
Poverty (2011): % below poverty line	15.0	9.8	12.3	25.3	27.6	27.03 (2007–2011)
Median usual weekly earnings of full-time wage and salary workers (25 yrs and older) (2011): $	$797	$825	$901	$582	$643	…
Less than a high school diploma	451	458	448	419	416	…
High school graduate, no college	638	663	564	568	538	…
Some college, no degree	719	743	710	643	611	…
Associate degree	768	95	713	706	624	…
Bachelor's degree and higher	1150	1165	1124	1000	958	…
Unemployed (2011): %	8.9	7.2	7.0	11.5	15.9	14.6
Incarceration of men (2008), per 100,000 population	1403	727	…	1,760	11,137	…
Political parity ratio (2004), defined as: (% in political office/% in population)						
—Congress	…	Men, 2.28; Women, 0.30	Men, 0.53; Women, 0.00	Men, 0.49; Women, 0.21	Men, 0.84; Women, 0.33	…
—State legislatures	…	Men, 2.04; Women, 0.52	Men, 0.41; Women, 0.15	Men, 0.34; Women, 0.49	Men, 0.89; Women, 0.46	…

(*continued*)

TABLE 3.3: Continued

Racial/Ethnic Inequity Indicators	Total US Population	White Non-Hispanic[a]	Asian[b]	Hispanic	Black Non-Hispanic[c]	American Indian/ Alaska Native
No health insurance, persons < 65 y (2011): %	17.2	12.9	16.5	31.1	18.8	34.2
Infant mortality rate (2008), per 1000 live births	6.6	5.5	4.5	5.6	12.7	8.4
Person-years lost per 100,000 population before 75 years of age (2010)	6642.9	6545.3	3061.2	4795.1	9832.5	6771.3
Self-reported fair or poor health status (2011): %	10.0	8.4	8.7	13.2	15.0	14.4
Severe psychological distress (2010–2011)[d]: %	3.3	3.2	1.7	4.0	3.7	5.6

Note. Ellipses indicate data not reported, which is indicative of imposed invisibility and is an informative social fact characterizing social context.

[a] Earnings by education for white only; all other data are for white non-Hispanic.

[b] Economic data and data on self-reported health and psychological distress are for Asians only; all other health data reported for Asians and Pacific Islanders combined.

[c] Net worth, poverty, and person-years of life lost before 75 years of age are reported for blacks only; all other data for the black non-Hispanic population.

[d] Serious psychological distress in past 30 days among adults aged 18 years and older, as measured by the Kessler 6 scale (range = 0–24; severe psychological distress: ≥ 13).

Sources: (a) wealth: US Census (179); (b) poverty: US Census (180) and Macartney et al. (181); (c) employment: US Bureau of Labor Statistics (182); (d) incarceration: National Center for Health Statistics (183); (e) political parity ratio: Hardy-Fanta et al. (184); (f) health and health insurance data: National Center for Health Statistics (108).

negative attitudes. Strikingly, then, data from the 1990 General Social Survey revealed that fully 75% of white Americans agreed that "black and Hispanic people are more likely than whites to prefer living on welfare" and a majority concurred that "black and Hispanic people are more likely than whites to be lazy, violence-prone, less intelligent, and less patriotic"; in 2008, data from this survey documented that the majority of white Americans (just over 50%) continued to believe that, compared to whites, black Americans were lazier, and 30% held that they were less intelligent (40). Moreover, in 2010, a national poll conducted right after the election of Barack Obama as the first black US president found that, despite the data shown in Table 3.3, fully 48% of the US white population agreed with the statement, "Today discrimination against Whites has become as big as a problem as discrimination against Blacks and other minorities," a statement also endorsed by 56% of Republicans and 62% of Tea Party adherents (47). By contrast, 70% of black respondents and 68% of Hispanics disagreed, as did 68% of Democrats (47). These are ugly social facts, with profound implications for not only our body politic but also the very bodies in which we live, love, rejoice, suffer, and die.

THEORIZING DISCRIMINATION AS A DETERMINANT OF HEALTH INEQUITIES

Before reviewing the contemporary evidence and methods used to investigate whether discrimination harms health, I start with explication of the theoretical framework I use to inform my critique. The theory I draw on is the ecosocial theory of disease distribution (4, 6, 48–50), which concerns who and what drives social inequalities in health.

USING ECOSOCIAL THEORY TO GUIDE RESEARCH ON DISCRIMINATION AND HEALTH INEQUITIES

A central focus of ecosocial theory is on how we literally biologically embody exposures arising from our societal and ecological context, thereby producing population rates and distributions of health. At issue are socially patterned exposure-induced pathogenic pathways, mediated by physiology, behavior, and gene expression, that affect the development, growth, regulation, and death of our body's biological systems, organs, and cells, culminating in disease, disability, and death. The contrast is to frameworks that treat causes of disease—and of group differences in biological characteristics and disease rates—as primarily innate, for example, as has long been argued for racial/ethnic inequities in health (51–53).

Indeed, integral to ecosocial theory is a painful awareness of the contested history of scientific ideas and practice, whereby eminent scientists, including in the health sciences, have been just as or more likely to develop and use scientific frameworks that justify, rather than question, discrimination and social inequality (4). Case examples particularly relevant to epidemiologic research on discrimination and health, about which reams have been written, include not only scientific racism and eugenics, but also scientific sexism and heterosexism (5, 35, 52–55).

Ecosocial theory accordingly requires explicit consideration of pathways of embodiment in relation to types and levels of exposure, the period and spatial expanse involved (i.e., spatiotemporal scale), and historical context, along with phenomena that affect susceptibility and resistance to exposure, ranging from micro (e.g., role of the gut microbiome in innate immunity) to macro (e.g., social organizing to challenge health inequities). Also core are issues of accountability (causal responsibility for) and agency (the power and ability to act) at every level, because they pertain not only to the magnitude of health inequities but also to how they are monitored, analyzed, and addressed. A critical knowledge of history is thus essential: the history of the exposures and outcomes under consideration, and the history of contending ways in which scientists have, in the context of their times, debated possible causal links (4). As with any scientific theory (56), the point is to frame and guide analysis of the phenomena of interest—in this case, population distributions of health, disease, disability, and well-being—and, as with any reflexive science, to generate knowledge relevant to altering the phenomena under study, in this case, the existence of health inequities (4).

Figure 3.1 illustrates the components of an ecosocial analysis as applied to the issue of racism and health (4, 8, 57). To guide both the research questions posed and the methods used, ecosocial theory posits, as shown on the left-hand side of the figure, that inequitable race relations simultaneously—and not sequentially (8, 58): *(1)* benefit the groups who claim racial superiority at the expense of those whom they deem intrinsically inferior, *(2)* racialize biology to produce and justify the very categories used to demarcate racial/ethnic groups, and *(3)* generate inequitable living and working conditions that, via embodiment, result in the biological expression of racism—and hence racial/ethnic health inequities (8, 51). A corollary is that there are many pathways, not just one, by which discrimination could harm health.

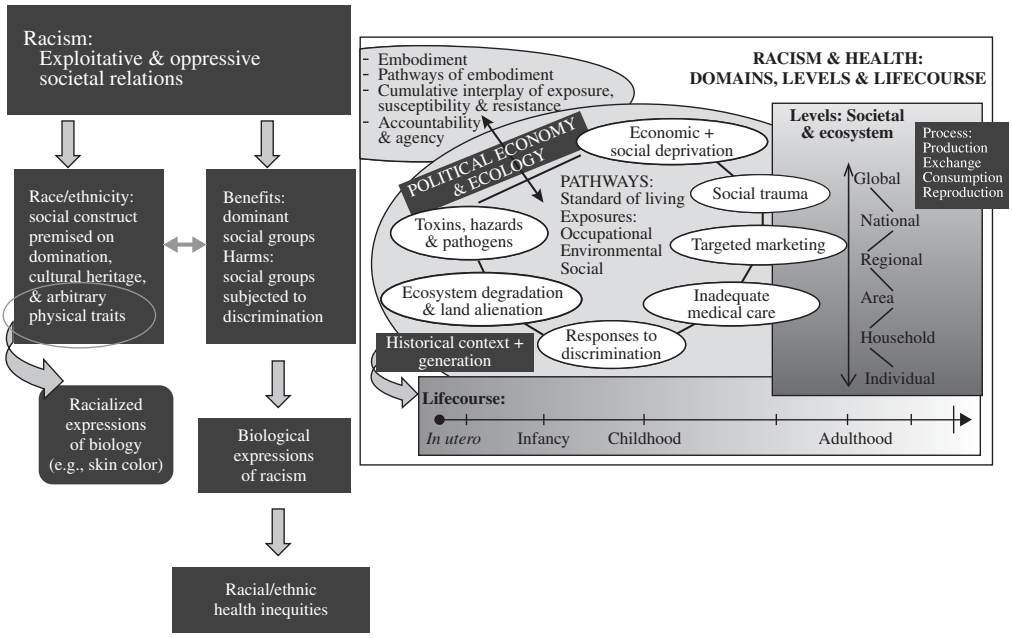


FIGURE 3.1: Ecosocial analysis of racism and health: core concepts and pathways of embodiment.
Sources: Krieger (1, 4, 8, 48, 57, 58).

The right-hand side of Figure 3.1 accordingly displays the major theorized pathways involved. They comprise: *(1)* economic and social deprivation; *(2)* excess exposure to toxins, hazards, and pathogens; *(3)* social trauma; *(4)* health-harming responses to discrimination; *(5)* targeted marketing of harmful commodities; *(6)* inadequate medical care; and *(7)* especially (but not only) for Indigenous peoples, ecosystem degradation and alienation from the land (1, 4, 8).

Moreover, as emphasized by ecosocial theory's simultaneous focus on exposure, susceptibility, and resistance, how people resist injustice and its health-harming effects, individually and collectively, and the resilience that enables them to do so also must be examined (1, 8, 48). Historical context in turn determines which pathways matter and are operative, at what level and at what point in the lifecourse. The implication is that just as expressions of discrimination can change, so too can their embodied manifestations, referring to both the kinds of health outcomes affected and the magnitude of the consequent and contingent health inequities.

The point is not that every study can or should attempt to measure every specified pathway at every level and at all relevant spatiotemporal scales. Rather, the value of a theoretical framework is that it can help concretize systematic substantive thinking about potential causal pathways, the constructs and entities employed and how they are operationalized and measured, the types of statistical analyses that should be conducted, potential threats to validity, and the complexities involved in interpreting study findings (4, 5, 56).

DISCRIMINATION AND STRUCTURED CHANCE

Thus, as ecosocial theory clarifies, to understand the impact of discrimination on population health, discrimination itself must be conceptualized as a dynamic population phenomenon that simultaneously structures individual risk and population rates of disease, thereby giving rise to health inequities. The population groups involved are not simply a priori entities, but instead active relational beings that arise from and are constituted by their underlying societal relationships (58). There can be no "black" without "white," no "Indigenous" without "colonizer," no "immigrant" without "native-born," no "women" without "men," no "LGBT" without "straight," no "disabled" without "abled," and no "elderly" without "young." The underlying causal presumption is that inequitable societal relationships between these co-defined societal groups shape each group's distribution of adverse exposures and health status, for good and for bad. By contrast, the causal reasoning of self-justifying discriminatory ideologies holds that characteristics of, distinctions between, and differential treatment of the groups at issue are "natural," due to innate differences (4, 5, 55).

One important corollary of the thesis that discrimination structures risk is that any random set of individuals selected from either of the groups, if subjected to the same adverse exposures, would have a greater likelihood of morbidity or mortality compared to individuals who are not exposed. The claim is not that each *person* would have the same elevated risk, given both the fundamental role of chance in disease etiology (59) and also heterogeneity among the unique individuals who constitute and are shaped by their membership in societal groups (58). Nor is the claim that all group differences in morbidity and mortality rates are inequities (60, 61). For example, only people who have prostates (aka men) can get prostate cancer, just as only people who have a cervix (aka women) can get cervical cancer, such that the sex-linked difference in rates of these cancers comparing men to women is strictly that: a difference—but, that said, among

men and among women there can be and are marked socioeconomic and racial/ethnic inequities in incidence, survival, and mortality (62). Consequently, where discrimination matters is for outcomes for which the on-average population risk would be similar across the societal groups at issue were it not for the adverse exposures due to structured differences in inequitable treatment.

Illustrating how discrimination can structure chance is an ingenious twist involving the first-ever mechanical device designed to provide a mechanical model of probability (Figure 3.2) (58). Its inventor was Sir Francis Galton (1822–1911), a highly influential British scientist and eugenicist, who himself coined the term "eugenics" and who held that heredity fundamentally trumped "environment" for traits influencing the capacity to thrive, whether physical, like health status, or mental, like "intelligence" (54, 63–65). In his 1889 opus *Natural Inheritance*, Galton sketched "an apparatus…that mimics in a very pretty way the conditions on which Deviation depends" (63: p. 63), whereby gunshot would be poured through a funnel down a board whose surface was studded with carefully placed pins, off which each pellet would ricochet, to be collected in evenly spaced bins at the bottom.

Galton termed his apparatus, which he apparently never built (65), the "Quincunx" because the pattern of the pins used to deflect the shot was like a tree-planting arrangement of that name, which at the time was popular among the English aristocracy (65). The essential point was that although each presumably identical ball had the same starting point, depending on the chance interplay of which pins it hit during its descent at which angle, it would end up in one or another bin. The accumulation of balls in any bin in turn would reflect the number of possible pathways (i.e., likelihood) leading to its ending up in that bin. Galton designed the pin pattern to yield a normal distribution. His conclusion?—that the device revealed (63: p. 66):

> …a wonderful form of cosmic order expressed by the "Law of Frequency of Error." The law would have been personified by the Greeks and deified, if they had known of it. It reigns

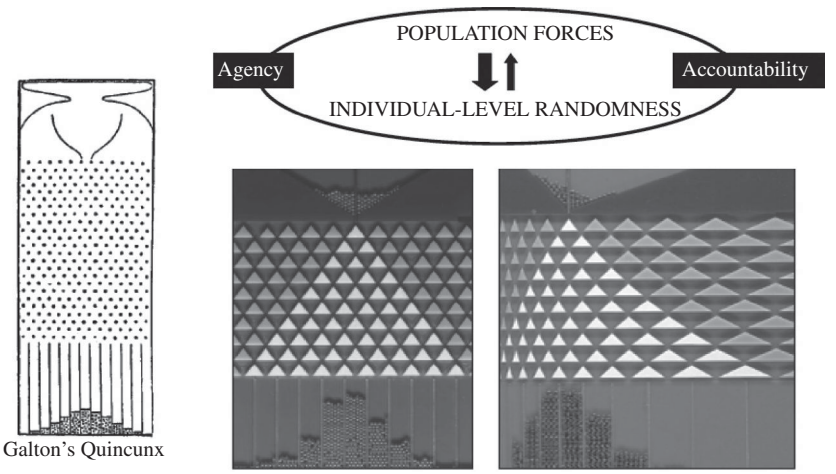


FIGURE 3.2: Producing population distributions: structured chance.

Sources: Krieger (58); Galton (63); Limpert et al. (66) (reproduced with permission).

> with serenity and in complete self-effacement amidst the wildest confusion. The huger the mob, and the greater the apparent anarchy the more perfect is its sway . . . each element, as it is sorted into place, finds, as it were, a pre-ordained niche, accurately adapted to fit it.

To Galton it was obvious that the observed distribution reflected the intrinsic properties of each "element" (in this case, the gun shot)—an inference that not only assumed the arrangement of the pins as a given but also conveniently sidestepped his role in designing their placement to produce a normal distribution.

However, a little more than a century later, some physicists not only built Galton's "Quincunx," as others have done (65), but went one further (66): they built two, one designed to generate the normal distribution and the other to generate the log-normal distribution (a type of distribution skewed on the normal scale, but for which the natural logarithm of the values displays a normal distribution) (Figure 3.2). As their devices clearly show, what structures the distribution is not the innate qualities of the "elements" themselves but the features of both the funnel and the pins—both their shape and placement. Together, these structural features determine which pellets can (or cannot) pass through the pins and, for those that do, their possible pathways.

The lesson is clear: altering the structure can change outcome probabilities, even for identical objects, thereby creating different population distributions. For the population sciences, this insight permits understanding how there can simultaneously be both chance variation *within* populations (individual risk) and patterned differences *between* population distributions (rates) (58, 59). Such an understanding of "structured chances" rejects explanations of population difference premised solely on determinism or chance and instead grounds the study of populations in real-life, historically contingent causal processes, including those structured by human agency, such as discrimination.

DISCRIMINATION AND HEALTH INEQUITIES: THE STATE OF THE EVIDENCE AND METHODS EMPLOYED

INDIRECT, DIRECT INDIVIDUAL, AND DIRECT INSTITUTIONAL MEASURES: INITIAL APPROACHES AND EVIDENCE LINKING DISCRIMINATION AND HEALTH (1980s–2000)

When I first reviewed the evidence on discrimination and health nearly 15 years ago, I delineated the three main approaches used to quantify health effects of discrimination (Figure 3.3). Listed in order of their frequency of use, from most to least common, they were (and remain): *(1)* indirectly, by inference, at the individual level; *(2)* directly, using measures of self-reported discrimination, at the individual level; and *(3)* in relation to institutional discrimination, at the population level. As I further explicated, all three approaches are informative, complementary, and necessary.

a. Indirect, at individual level: examine whether "known risk factors" explain differences in health outcomes between members of dominant and subordinated groups; if not, infer discrimination may contribute to residual difference				
Discrimination by physician (unobserved)	→ Differences in treatment (observed) Possibly affected by: – severity in illness – comorbidity – age – insurance status – economic resources – family support – patient "preference" (usually unobserved) etc.		→ Differences in outcome (observed)	
b. Direct, at individual-level: among subordinated group, examine whether self-reported experiences of discrimination are associated with specified health outcome (note: could also use implicit measure of exposure, or use vignette in laboratory setting to manipulate exposure)				
Discrimination (self-reported)	→ threat	→ fear anger denial etc.	→ physiologic responses – cardiovascular – endocrine – neurologic – immune etc.	→ health outcome (observed)
c. Institutional, at population-level: among subordinated group, examine whether population-level measures of discrimination are associated with population rates of health outcome (or, if data available, can analyze as multi-level model, and estimate impact of population exposure on individual risk)				
Discrimination (unobserved)	→ Residential segregation (observed)	→ Concentration of poverty, poor housing quality, increased population density, toxic exposures, lack of access to services and goods, political disempowerment, etc.		→ elevated morbidity and mortality rates (observed)

FIGURE 3.3: Three main epidemiologic approaches to studying the health effects of discrimination.
Source: Krieger (1).

In brief, for the "indirect" approach investigators compare health outcomes of subordinated and dominant groups, albeit without any direct data on exposure to discrimination. If distributions of these outcomes differ, then researchers determine whether observed disparities can be explained by "known risk factors." If so, investigators interpret their findings in light of how discrimination may shape distribution of the relevant "risk factors." If, however, a residual difference persists, even after controlling for these other risk factors, then additional aspects of discrimination may be inferred as a possible explanation for the remaining disparities (assuming no unmeasured confounders and also no differential bias in measurement of the relevant "risk factors").

Although the weaknesses of making causal inferences based on omitted variables is well known, two factors spur use of the "indirect" approach. The first is the overwhelming lack of data

on discrimination, either interpersonal or institutional, in most leading sources of population data (12), including for population health, for example, vital records, cancer registries, national surveys, and the vast majority of epidemiologic studies designed without considering whether discrimination might affect the health outcomes of interest (1). Partially accounting for this omission, beyond ideological aversion (45), is skepticism as to the feasibility of obtaining valid data on discrimination (1, 12). The indirect approach nevertheless has been used because, from a "lesser of two evils" standpoint, it is arguably better to determine whether "known risk factors" (especially economic) can—or cannot—account for observed societal health inequities, even absent data on discrimination, as opposed to leaving the question unasked and the inequities unanalyzed.

The second reason for use of the indirect approach is more affirmative and points to one of its cardinal strengths: its utility for analyzing outcomes where determination of whether discrimination has occurred requires an individual knowing about not only her or his own experience but also that of others (1, 12). A case in point is biased medical decisions by health providers; in such circumstances, indirect statistical evidence of discrimination based on medical records is the only available option, short of conducting studies that focus on the health care providers directly (as opposed to the people they treat) (1, 12, 23).

Attesting to the utility of the "indirect" approach, robust epidemiologic evidence indicates that racial/ethnic inequities in current and cumulative impoverishment versus affluence contribute substantially to explaining racial/ethnic inequities in disease occurrence, survival, and mortality, with examples of this work stretching from the mid-nineteenth century CE to the present day (67–74). Additional historical, sociological, and economic research (both observational and experimental) in turn provides robust evidence that institutional and interpersonal discrimination, present and past, contribute to contemporary racial/ethnic inequities in income, wealth, and education (12, 18, 31, 75, 76). Accordingly, as I previously noted, studies using indirect approaches to measuring health effects of discrimination can and do provide essential, powerful, and important evidence that discrimination shapes societal distributions of health and disease. Nevertheless, as I also observed, epidemiologic studies using the "indirect" approach remain vulnerable to: *(1)* nondifferential and differential measurement error (e.g., for socioeconomic position and other "risk factors" included), and *(2)* contending explanations of any observed residual difference (e.g., in the case of racial/ethnic health inequities, whether the remaining differences are due to racism versus "race," conceptualized as innate racial difference) (1, 77).

Consequently, to meet the challenge of explicitly measuring people's direct experiences of discrimination and relating this to their health status, starting in the 1980s a new generation of public health researchers began devising new methods and approaches (1). Skeptical instead of the common view that rigorous discrimination measures were impossible to devise, the emphasis was and remains on development of valid and reliable instruments to measure individuals' exposure to discrimination across the lifecourse, whether as direct target or as witness, and also their responses to this exposure: behaviorally, psychologically, and physiologically. One gap impelling this work was the absence of any well-characterized, "ready-to-use," validated instruments in the social science literature that were appropriate for large-scale empirical studies. Instead, most empirical sociologic studies on discrimination at the time had either focused chiefly on racial attitudes of people who discriminate, rather than experiences of those who have endured discrimination (40, 46, 78), or else, as was also the case in psychological research, they employed in-depth interviews and other qualitative approaches not readily transferable to epidemiologic research (30, 79–81). Second, the

measures developed in these kindred disciplines understandably were not concerned with measuring exposure in relation to the health-specific concern of etiologic period, that is, time between exposure, onset of pathogenic processes, and occurrence of disease, nor were they engaged with how biological effects could be different depending on the point in the lifecourse when someone was exposed, from in utero onward (1, 8, 82, 83).

With regard to outcomes analyzed in relation to self-reported exposure to discrimination, in these initial studies the most common by far pertained to mental health, for example, depression, psychological distress, the second most frequent was hypertension or blood pressure, and other outcomes considered (but not by more than one study) included smoking and inadequate health care. Overall, these studies consistently reported that higher levels of self-reported experiences of discrimination were associated with poorer mental health; associations with somatic health were both more complex and inconsistent.

Moreover, extending the idea of direct measurement to experimental design, two of the early studies reported that blood pressure and heart rate among African Americans increased more quickly upon viewing movie scenes or imagining scenarios involving racist, as compared to nonracist but angry, or neutral, encounters (84, 85). Then as now (12), the experimental studies were conceptualized as having both strengths and limitations. Strengths included their ability to control exposure, to test particular biological pathways, and to use randomization to avoid or minimize concerns about the potential confounding endemic to observational studies. Recognized limitations of lab-based studies were that, by design, they could (1) only manipulate short-term psychosocial exposures (and hence not any of the other types of pathways potentially implicated in how discrimination harms health, for example, chronic economic deprivation and social exclusion), and (2) only observe acute responses, whose predictive value for disease pathogenesis often is debatable; a third concern pertains to generalizability, depending on the type of selection bias involved in recruiting persons to be part of lab-based studies (12).

The third approach, whose use began in earnest in the early 1990s, shifted the focus to exposures that can be measured only at the population level, most typically in relation to residential segregation and also regarding population-level expressions of empowerment, such as representation in government. Early studies using this third strategy provided evidence that African American morbidity and mortality rates were associated with not only residential segregation (building on prior work by DuBois [69] and also Yankauer [86]), but also racial/ethnic political clout and regional racial attitudes (87–91).

Three types of spatiotemporal threats to validity, however, recognized at the time, tempered interpretation of results (1). One concerned aggregation bias, since the initial wave of studies relied heavily on group-level measures of both exposures and outcomes. The now common use of multilevel analysis (92, 93) has at least methodologically addressed this potential threat to validity, with the challenge to researchers now shifted to assembling data bases with the relevant individual-level and contextual data. Also at issue was assessment of exposure in relation to not only etiologic period but also residential mobility, as extensively discussed in the sociological literature (12). Thus, from a temporal standpoint, an association of higher levels of residential segregation or negative racial attitudes with, say, concurrent infant mortality rates or childhood morbidity rates or homicide rates would provide more compelling evidence of health effects of segregation or racial attitudes than would its association with all-cause mortality among adults, given the much longer latency period for most causes of death (e.g., cardiovascular disease, cancer). If, however, current

levels of segregation reflected past levels and little bias were introduced by residential mobility, the threat to causal inference would be lessened but not eliminated (1).

In summary, the first wave of explicit research on discrimination and health focused chiefly on racial discrimination and provided provocative evidence of diverse pathogenic pathways: via limiting access to educational, economic, occupational, residential, and political resources, thereby constraining options for living and working in healthy environments, and via serving as a stressor that adversely affected both psychological well-being and health behaviors, thereby increasing risk of somatic and mental illness. Study limitations notwithstanding, the net effect was to support the hypothesis that discrimination could elevate risk of adverse mental and somatic outcomes.

THE RESEARCH TODAY: A REVIEW OF REVIEW ARTICLES (AS OF MID-2013)

By sheer count of the hundreds of empirical investigations cataloged in contemporary review articles that explicitly focus on discrimination and health (Table 3.4; see citations and search strategy in the table's footnote), research on this topic has dramatically escalated in the early twenty-first century CE—at least for some types discrimination, not others. Racial discrimination, for example, has been the focus of 40 review articles, all but 3 published during or after 2000, that together analyze well over 350 different studies—and although most research remains US based, with a focus especially on African Americans, a growing number of studies pertain to immigrants of color and to Indigenous Peoples, and studies now hail from diverse countries in Latin America and Europe and also from Canada, New Zealand, Australia, and Japan. Research focused on discrimination against lesbian, gay, bisexual, and transgender (LGBT) persons has likewise grown considerably, with the 9 published review articles, all appearing during or after 2000, encompassing over 50 empirical investigations, predominantly from the United States, but also including Canada, Australia, and several European countries.

In the case of gender discrimination and health, the 10 identified review articles (all but one published during or after 2000) analyzed upward of 250 articles, mainly from the United States, of which 80% focused on gender bias in medical care. Of note, search strategies focused on gender discrimination, bias, and prejudice yielded scant review articles pertaining to either disease etiology or to health in relation to self-reported experiences of gender discrimination, a finding likely reflecting the growing tendency in public health and medicine to reframe analysis of social inequalities in women's health in relation to "gender roles" and "gender-based violence" (94–97), absent use of terminology regarding—or instruments to measure—exposure to gender discrimination. Inclusion of epidemiologic review articles focused on violence against women and sexual abuse (including of children) not also indexed by terms pertaining to discrimination would have added another 60+ review articles.

By contrast, the number of review articles—and empirical studies—explicitly focused on discrimination and health in relation to disability and age remains small: 4 and 3 respectively, all of which acknowledged the scant evidence available (apart from studies on physical and sexual abuse). In both cases, the primary emphasis, as with the research on gender, pertained to discrimination in the health care system, even though most of the review articles acknowledged that discrimination in employment against disabled and older adults can affect

TABLE 3.4: Empirical review articles on discrimination and health by type of discrimination: January 1, 1900–June 1, 2013

Author & Year	Characteristics of Empirical Articles Reviewed on Discrimination and Health		Key Findings
1) Multiple types (n = 5)			
Krieger et al. (71), 1993	**Health outcomes**	P, M, HB, HC	pp. 101–102: "In addition to recognizing the profound significance of social class in shaping black/white differences in disease occurrence, several studies are beginning to explore some of the other adverse effects of other forms of racism on the health of black Americans, including outright discrimination within each socioeconomic stratum . . . ."; pp. 104–105: "Much of the new ferment regarding 'women's health' and the health of women is primarily concerned with sexism in both medicine and biomedical research. . . . Outside of medical interventions, however, little epidemiologic research has directly examined how non-violent forms of sexism (i.e., other than domestic violence, rape, etc.) may affect women's health . . . ."; p. 108: "Provocative research on the 'hidden injuries of class,' moreover, suggest that discrimination, exclusion, and subordination based on class might produce the same types of psychosocial trauma as racism and sexism."
	N of articles	10	
	Years	1984–1991	
	Countries	US	
	Study designs	O (9), E (1)	
Krieger (1), 1999/2000	**Health outcomes**	P, M, HB	p. 311 (1999): "The most common outcome (ten studies) was mental ill-health, such as depression, psychological distress; the second most common (five studies) was hypertension or blood pressure. Overall, studies consistently reported that higher levels of self-reported experience of discrimination were associated with poorer mental health; associations with somatic health . . . were more complex."
	N of articles	20	
	Years	1984–1999	
	Countries	US	
	Study designs	O	

Pascoe and Richman (185), 2009	**Health outcomes**	P, M, HB, stress response	p. 531: "Analysis of 134 samples suggests that when weighting each study's contribution by sample size, perceived discrimination has a significant negative effect on both mental and physical health. Perceived discrimination also produces significantly heightened stress responses and is related to participation in unhealthy and nonparticipation in healthy behaviors."
	N of articles	134 (meta-analysis)	
	Years	1986–2007	
	Countries	US, NZ	
	Study designs	O (119); E (15)	
Santry and Wren (186), 2012	**Health outcomes**	HC	p. 137: "Racial, ethnic, and gender disparities in health outcomes are a major challenge for the US health care system. Although the causes of these disparities are multifactorial, unconscious bias on the part of health care providers plays a role. Unconscious bias occurs when subconscious prejudicial beliefs about stereotypical individual attributes result in an automatic and unconscious reaction and/or behavior based on those beliefs. This article reviews the evidence in support of unconscious bias and resultant disparate health outcomes"
	N of articles	63	
	Years	1991–2011	
	Countries	US	
	Study designs	O (50), E (13)	
Goto et al. (187), 2013	**Health outcomes**	M	p. 445: "Positive and statistically significant associations were observed between discrimination and adverse mental health outcomes, particularly substance use, depression, and alcohol-related disorders. Only one third of the studies explicitly adopted a specific theoretical framework to interpret the examined relationships."
	N of articles	34	
	Years	2000–2010	
	Countries	US, Mexico, Holland	
	Study designs	O	

(*continued*)

TABLE 3.4: Continued

Author & Year	Characteristics of Empirical Articles Reviewed on Discrimination and Health		Key Findings
2) Racial discrimination, including Indigenous people and immigrants of color (n = 40)			
a) Adult persons of color (n = 16)			
—see also "multiple types": Krieger et al. (71), 8 studies; Krieger (1), 15 studies; Pascoe and Richman (185), 125 studies; Santry and Wren (186), 53 studies; Goto et al. (187), 27 studies			
Williams and Collins (72), 1995	**Health outcomes**	P, M, HB, HC	pp. 366–367: "Racism can affect health in at least three ways. First, it can transform social status so that SES indicators are not equivalent across race.... Second, racism can restrict access to the quantity and quality of health-related desirable services such as public education, health care, housing, and recreational facilities....Third, the experience of racial discrimination and other forms of racism may induce psychological distress that may adversely affect physical and mental health status, as well as the likelihood of engaging in violence and addiction."
	N of articles	6	
	Years	1984–1993	
	Countries	US	
	Study designs	O	
Williams and Williams-Morris (188), 2000	**Health outcomes**	M	p. 243: "...the available scientific evidence suggests racism can adversely affect mental health status in three ways. First racism in societal institutions can lead to truncated socioeconomic mobility, differential access to desirable resources, and poor living conditions that can adversely affect mental health. Second, experiences of discrimination can induce physiological and psychological reactions that can lead to adverse changes in mental health status. Third, in race-conscious societies, the acceptance of negative cultural stereotypes can lead to unfavorable self-evaluations that have deleterious effects on psychological well-being."
	N of articles	15	
	Years	1987–2000	
	Countries	US, Canada	
	Study designs	O (community-based)	

Williams et al. (189), 2003	**Health outcomes**	P, M, HB	p. 213: "The authors review the available empirical evidence from population-based studies of the association between perceptions of racial/ethnic discrimination and health. This research indicates that discrimination is associated with multiple indicators of poorer physical and, especially, mental health status.... Research on stress points to important directions for the future assessment of discrimination and the testing of the underlying processes and mechanisms by which discrimination can lead to changes in health."
	N of articles	53	
	Years	1988–2002	
	Countries	US, UK	
	Study designs	O (community-based)	
Schnittker and McLeod (190), 2005	**Health outcomes**	P, M, HB	pp. 89–90: "Self-reported discrimination is significantly associated with physical and mental health" including "major depression," "generalized anxiety disorders," and "self-rated health, chronic conditions, disability, and blood pressure and other cardiovascular risk factors, although the results for the latter are complex."
	N of articles	17	
	Years	1990–2003	
	Countries	US	
	Study designs	O	
Paradies (191), 2006	**Health outcomes**	P, M, HB	p. 888: "This paper reviews 138 empirical quantitative population-based studies of self-reported racism and health. These studies show an association between self-reported racism and ill health for oppressed racial groups after adjustment for a range of confounders. The strongest and most consistent findings are for negative mental health outcomes and health-related behaviours, with weaker associations existing for positive mental health outcomes, self-assessed health status, and physical health outcomes."
	N of articles	138	
	Years	1980–2004	
	Countries	US primarily; also: Canada, Australia, NZ Barbados, Dominica	
	Study designs	O (population-based)	

(*continued*)

TABLE 3.4: Continued

Author & Year	Characteristics of Empirical Articles Reviewed on Discrimination and Health		Key Findings
Williams and Mohammed (192), 2009	**Health outcomes**	P, M, HB, HC	p. 22: "Studies of mental health continue to dominate the discrimination and health literature." p. 39: "The research on discrimination and health is continuing to grow rapidly. Although the discrimination variable has been operationalized in a variety of ways, the consistency of an inverse association between discrimination for an increasingly broad range of health outcomes, across multiple population groups in a wide range of cultural and national contexts is impressive, and lends credibility to the plausibility of perceived discrimination as an important emerging risk factor for disease."
	N of articles	115	
	Years	2005–2007	
	Countries	US, Canada, UK, Denmark, Sweden, Nether-lands, Austria, Bosnia, Croatia, S Africa, NZ, Australia	
	Study designs	O (106); E (9)	
Brondolo et al. (193), 2009	**Health outcomes**	P, M	p. 74: "The effects of racial identity on mental and physical health are complex, and the data do not support a uniformly positive effect of each aspect of racial or ethnic identity on mental health"; p. 76: "Overall, the quantitative literature provides minimal support for the hypothesis that social support . . . buffers the impact of racism on psychological health . . . [or] on indices of physical health"; pp. 83–83: "The psychobiological effects of anger suppression among African Americans are among the most consistent findings in the literature on coping with racism. These data suggest that suppressing anger in the face of discrimination is associated with elevated BP or greater BP responses."
	N of articles	24	
	Years	1989–2007	
	Countries	US	
	Study designs	O (20); E (4)	

Brondolo et al. (194), 2011	**Health outcomes**	P	p. 518: "Direct evidence linking individual/interpersonal racism to [hypertension] HTN diagnosis is weak. However, the relationship of individual/interpersonal racism to ambulatory blood pressure (ABP) is more consistent, with all published studies reporting a positive relationship of interpersonal racism to ABP. There is no direct evidence linking internalized racism to BP. Population-based studies provide some evidence linking institutional racism, in the forms of residential racial segregation (RRS) and incarceration, to HTN incidence. Racism shows associations to stress exposure and reactivity as well as associations to established HTN-related risk factors including obesity, low levels of physical activity and alcohol use. The effects vary by level of racism."
	N of articles	24	
	Years	1984–2010	
	Countries	US	
	Study designs	O	
Brondolo et al. (195), 2012	**Health outcomes**	P, M, HB	p. 359: "The literature suggests that cultural, institutional, interpersonal and internalized racism affect peer relationships through a variety of pathways. In turn, racism-related effects on the development of peer relationships contribute to racial disparities in economic outcomes and health status."
	N of articles	9	
	Years	2002–2009	
	Countries	US	
	Study designs	O	
Shavers et al. (147), 2012	**Health outcomes**	HC	p. 962: "We examined the availability of data on the prevalence, trends, mechanisms, and institutional policies and practices associated with racial/ethnic discrimination in health care settings. Although there were a number of studies that described race/ethnicity based discriminatory behaviors, attitudes, biases and preferences that could potentially contribute to discriminatory health care we found no studies that specifically addressed the US prevalence or trends. Also, relatively absent were studies that addressed how institutional racism impacts the health care received by racial/ethnic minority patients."
	N of articles	58	
	Years	2008–2011	
	Countries	US	
	Study designs	O; E	

(*continued*)

TABLE 3.4: Continued

Author & Year	Characteristics of Empirical Articles Reviewed on Discrimination and Health		Key Findings
Couto et al. (196), 2012	**Health outcomes**	P	p. 956: "The objective of this study was to update previous literature reviews on discrimination and blood pressure . . . In the 22 studies assessed, the association between discrimination and blood pressure/hypertension was assessed 50 times. Twenty results (40%) showed no association between them, and only 15 (30%) revealed global positive associations, of which 67% were statistically significant. Eight negative associations were also observed, suggesting that higher exposure to discrimination would be associated with lower blood pressure/hypertension."
	N of articles	22	
	Years	2000–2010	
	Countries	US	
	Study designs	O	
b) children of color (n = 2)			
Pachter and Coll (197), 2009	**Health outcomes**	P, M, HB	p. 260: "The studies in this review show that that racism is associated with differential health outcomes and can adversely affect psychological and physiological functioning. Furthermore, they provide data to show that racism is associated with negative biopsychosocial sequelae that may contribute to health disparities, and different exposure to it contributes to variability in health outcomes within groups. This is most evident in the literature demonstrating its association with poor birth outcomes."
	N of articles	49	
	Years	1994–2007	
	Countries	US	
	Study designs	O	
Sanders-Phillips et al. (198), 2009	**Health outcomes**	P, M, HB	p. S176: ". . . the social environment for many children of color includes personal and family experiences of racial discrimination that foster perceptions of powerlessness, inequality, and injustice. In turn, these perceptions may influence child health outcomes and disparities by affecting biological functioning (eg, cardiovascular and immune function) and the quality of the parent-child relationship and promoting psychological distress (eg, self-efficacy, depression, anger) that can be associated with risk-taking and unhealthy behaviors."
	N of articles	20	
	Years	1997–2008	
	Countries	US	
	Study designs	O	

c) African American or black (only) (n = 7)			
Wyatt et al. (199), 2003	**Health outcomes**	Cardiovascular	p. 327: "Despite hypothesized links between perceptions of racism, perceived discrimination, and CVD, few population-based studies have examined these associations. Findings, though inconsistent, do suggest the potential for such a relationship, particularly with hypertension."
	N of articles	19	
	Years	1984–2003	
	Countries	US, UK, Finland	
	Study designs	O	
Brondolo et al. (200), 2003	**Health outcomes**	Cardiovascular	p. 62: "The existing data on the relationship of racism to BP [blood pressure] level or HT [hypertension] status are mixed. These pioneering studies provide important insights and guidance, but methodological limitations limit their interpretability and are likely to account for the inconsistent and relatively weak findings. A comprehensive evaluation of the relationship between racism and BP status has yet to be conducted. Findings from CVR [cardiovascular reactivity] studies are clearer. These studies suggest that acute exposure to racism is associated with increases in cardiovascular activation. In addition, past exposure to racism may influence current CVR to race-related and other stressors."
	N of articles	17	
	Years	1972–2001	
	Countries	US	
	Study designs	O (6); E (11)	
Giscombé and Lobel (201), 2005	**Health outcomes**	Adverse birth outcomes	p. 662: "The authors examine 5 explanations for these differences in rates of adverse birth outcomes: (a) ethnic differences in health behaviors and socioeconomic status; (b) higher levels of stress in African American women; (c) greater susceptibility to stress in African Americans; (d) the impact of racism acting either as a contributor to stress or as a factor that exacerbates stress effects; and (e) ethnic differences in stress-related neuroendocrine, vascular, and immunological processes. The review of literature indicates that each explanation has some merit, although none is sufficient to explain ethnic disparities in adverse birth outcomes. There is a lack of studies examining the impact of such factors jointly and interactively."
	N of articles	6	
	Years	1996–2005	
	Countries	US	
	Study designs	O	

(*continued*)

TABLE 3.4: Continued

Author & Year	Characteristics of Empirical Articles Reviewed on Discrimination and Health		Key Findings
Mays et al. (202), 2007	**Health outcomes**	P, M, HB	p. 201: "In this article, we review emerging work that locates the cause of race-based health disparities in the external effects of the contextual social space on the internal world of brain functioning and physiologic response. These approaches reflect the growing interdisciplinary nature of psychology in general, and the field of race relations in particular."
	N of articles	27	
	Years	1989–2006	
	Countries	US	
	Study designs	O, E	
Giurgescu et al. (203), 2011	**Health outcomes**	Adverse birth outcomes	p. 362: "A consistent positive relationship existed between perceptions of racial discrimination and preterm birth, low birth weight, and very low birth weight. No relationship was found between racial discrimination and gestational age at birth."
	N of articles	10	
	Years	1996–2009	
	Countries	US	
	Study designs	O	
Pieterse et al. (204), 2012	**Health outcomes**	M	pp. 5–6: "...the aggregated correlation between perceived racism and psychological distress for 66 studies using a random-effects model was r = .20, 95% CI [0.17, 0.22]... results provided graphical and statistical support that publication bias was not present... the lack of significant moderation findings for sample type, publication type, and racism scale type suggest the relationship between perceived racism and mental health for Black Americans is quite robust."
	N of articles	66 (meta-analysis)	
	Years	1966–2011	
	Countries	US	
	Study designs	O	

Cuffee et al. (205), 2012	**Health outcomes**	Cardiovascular	p. 422: "This systematic review supports the association of racial discrimination with an increased risk of developing hypertension; however, the picture is not uniform. Methodological challenges, such as floor or ceiling effects of reported discrimination and low sample size, may have prevented researchers from detecting important associations."
	N of articles	15	
	Years	1990–2010	
	Countries	US	
	Study designs	O	
d) Indigenous peoples (n = 2)			
Paradies (206), 2006	**Health outcomes**	P, M, HB	—focused on both Indigenous Peoples (n = 17 studies) and African Americans (n = 48 studies); p. 295: "A range of chronic diseases, as well as harmful health behaviors, were associated with psychosocial stress for indigenous peoples and African Americans, with much stronger findings for mental rather than physical health outcomes."
	N of articles	65 (Indigenous: 17)	
	Years	1973–2004	
	Countries	US, Australia, NZ	
	Study designs	O (50); intervention (15)	
Walters et al. (124), 2011	**Health outcomes**	P, M, HB, + spiritual	—focused on historical trauma among American Indians and Alaska Natives (AIAN); p. 185: "a high proportion of AIANs have high levels of historical-trauma loss manifesting in thinking about the impact of land-based trauma, on a weekly, and in some cases, daily basis. Moreover, after controlling for contemporary trauma, we found that HT land-based events continued to have a significant effect on mental and physical health"
	N of articles	3	
	Years	1999–2011	
	Countries	US	
	Study designs	O	

(*continued*)

TABLE 3.4: Continued

Author & Year	Characteristics of Empirical Articles Reviewed on Discrimination and Health		Key Findings
e) Immigrants (Asian and Latinos, solely immigrants and for US also including both immigrant and US-born) (n = 4)			
Gee et al. (166), 2009	**Health outcomes**	P, M, HB	—focused on Asian immigrants; p. 130: "The majority of articles focused on mental health problems, followed by physical and behavioral problems. Most studies find that discrimination was associated with poorer health, although the most consistent findings were for mental health problems."
	N of articles	62	
	Years	1960–2009	
	Countries	US, UK, Canada, NZ, Australia, Finland, Japan	
	Study designs	O	
Nadimpalli and Hutchinson (207), 2012	**Health outcomes**	P, M	—focused on Asian Americans; p. 127: "Discrimination was associated with depressive symptoms in seven studies. Three studies found associations between discrimination and physical health, including cardiovascular disease, respiratory conditions, obesity, and diabetes. Although the literature was limited by self-reported data, cross-sectional designs, and inconsistent definitions and measurement of discrimination, the findings suggest that discrimination is a significant contributor to poorer health and health disparities for Asian Americans."
	N of articles	14	
	Years	2002–2011	
	Countries	US	
	Study designs	O	
Clough et al. (208), 2013	**Health outcomes**	HC	—focused on US Asian immigrants; p. 387: "Four main themes emerged from the literature with respect to health care access and quality among Asian immigrants: access to health services; linguistic discordance and health communication between patient and provider; health-related beliefs of patients and cultural incompetency of health systems; and perceived discrimination in the health care setting."
	N of articles	7	
	Years	1980–2011	
	Countries	US	
	Study designs	O	

Viruell-Fuentes et al. (167), 2012	**Health outcomes**	P, M, HB	—focused on US immigrants; p. 2100–1: "...even though some results are mixed, there is growing evidence that perceived discrimination is associated with lower levels of physical and mental health; poor access to quality health care; and certain deleterious health behaviors...the strength of the association between discrimination and health among immigrants appears to vary both by length of time in the United States and age at migration"; p. 2102: "for some, such as U.S.-born racialized groups, living in immigrant/ethnic enclaves may reflect limited opportunities for social, economic, and residential mobility...in some cases ethnic/immigrant enclaves in traditional urban immigrant gateway areas may provide a supportive environment for immigrants"; p. 2013: "Although research on the health effects of immigration policies is sparse, several studies point to their importance for health"
	N of articles	14	
	Years	2000–2011	
	Countries	US	
	Study designs	O	
f) Residential segregation and environmental racism (n = 7)			
Acevedo-Garcia et al. (209), 2003	**Health outcomes**	P, HB, env exposures	p. 216: "The majority of studies investigated the effect of racial residential segregation on the health of African Americans...a majority of the studies revealed a detrimental effect of Black-White segregation on African American mortality outcomes on Black-White mortality differentials..."
	N of articles	29	
	Years	1966–2002	
	Countries	US	
	Study designs	O	
Brulle and Pellow (210), 2006	**Health outcomes**	P, env exposures	pp. 103–104: "...In many communities, it is people of color and the poor who tend to live near environmentally hazardous facilities and who bear a larger share of the health burden from exposure to toxins..."
	N of articles	13 (population-based)	
	Years	1972–2005	
	Countries	US	
	Study designs	O	

(*continued*)

TABLE 3.4: Continued

Author & Year	Characteristics of Empirical Articles Reviewed on Discrimination and Health		Key Findings
Mohai et al. (211), 2009	**Health outcomes**	P, env exposures	p. 406: "Today, hundreds of studies conclude that, in general, ethnic minorities, indigenous persons, people of color, and low-income communities confront a higher burden of environmental exposure from air, water, and soil pollution from industrialization, militarization, and consumer practices. Known variously as environmental racism, environmental inequality, or environmental injustice, this phenomenon has also captured the attention of policy makers."
	N of articles	40 (population-based)	
	Years	1970–2007	
	Countries	US	
	Study designs	O	
Kramer and Hogue (212), 2009	**Health outcomes**	P	p. 178: "Thirty-nine identified studies test an association between segregation and health outcomes. The health effects of segregation are relatively consistent, but complex. Isolation segregation is associated with poor pregnancy outcomes and increased mortality for blacks, but several studies report health-protective effects of living in clustered black neighborhoods net of social and economic isolation. The majority of reviewed studies are cross-sectional and use coarse measures of segregation."
	N of articles	39	
	Years	1950–2008	
	Countries	US	
	Study designs	O	
Landrine and Corral (213), 2009	**Health outcomes**	P, HB, HC	p. 183: "Black and White neighborhoods are separate and unequal in a manner that may explain the separate, unequal health profiles of Blacks and Whites—particularly because most Blacks reside in mostly Black neighborhoods irrespective of their individual-level SES."
	N of articles	31	
	Years	2000–2008	
	Countries	US	
	Study designs	O	
White and Borrell (214), 2011	**Health outcomes**	P, M, HB	p. 441: "The majority of the reported findings to date identify an association between residence in highly segregated areas and deleterious health outcomes, and to a lesser extent suggest a protective effect of segregation. The literature is fairly limited with regard to examining adult chronic diseases and health behaviors."
	N of articles	45	
	Years	1950–2009	
	Countries	US	
	Study designs	O	

White et al. (215), 2012	**Health outcomes**	HC	p. 1278: "Increasing evidence indicates that racial/ethnic residential segregation is a key factor driving place-based health care inequalities"
	N of articles	13	
	Years	1998–2011	
	Countries	US	
	Study designs	O	
g) Methodological (instruments to measure self-reported experiences of racial discrimination) (n = 2)			
Kressin et al. (216), 2008	**Health outcomes**	HC	p. 697: "We identified 34 measures of racism/discrimination; 16 specifically assessed dynamics in the health care setting. Few measures were theoretically based; most assessed only general dimensions of racism and focused specifically on the experiences of African American patients. Acceptable psychometric properties were documented for about half of the instruments."
	N of articles	34 measures	
	Years	1986–2007	
	Countries	US	
	Study designs	Psychometric	
Bastos et al. (217), 2010	**Health outcomes**	all	p. 1091: "Despite the fact that discrimination stands as topic of international relevance, 23 (96%) scales were developed within the United States. Most studies (67%, N = 16) were published in the last 12 years, documenting initial attempts at scale development, with a dearth of investigations on scale refinements or cross-cultural adaptations. Psychometric properties were acceptable; sixteen of all scales presented reliability scores above 0.7, 19 out of 20 instruments confirmed at least 75% of all previously stated hypotheses regarding the constructs under consideration, and conceptual dimensional structure was supported by means of any type of factor analysis in 17 of 21 scales."
	N of articles	24 scales	
	Years	1973–2008	
	Countries	US (except 1 Australian)	
	Study designs	Psychometric	

(*continued*)

TABLE 3.4: Continued

Author & Year	Characteristics of Empirical Articles Reviewed on Discrimination and Health		Key Findings
3) Gender (n = 10)			
—see also "multiple types": Krieger et al. (71), 2 studies; Krieger (1), 3 studies; Pascoe and Richman (185), 13 studies; Santry and Wren (186), 11 studies; Goto et al. (187), 4 studies			
Swanson (218), 2000	**Health outcomes**	P, M, HB	p. 77: "Sexual harassment—unwanted, unsolicited verbal or physical behaviors—can be a severe occupational stressor with serious physical, psychological, behavioral and career consequences. . . .Targets have reported a range of psychological symptoms such as depression, anxiety, fearfulness, feelings of guilt and shame; such physical symptoms as headaches, gastrointestinal disorders, and sleep disorders; and such job-related outcomes as job withdrawal, negative job attitudes, involuntary job loss, and career interruption. Several studies have shown that sexual harassment is a particularly noxious stressor for women and has a significant impact on psychological distress and absenteeism beyond that attributable to regular job stressors."
	N of articles	7	
	Years	1986–1997	
	Countries	US	
	Study designs	O	
Raine (219), 2000	**Health outcomes**	HC	p. 237: "One hundred and thirty-eight studies were identified covering five major topics: coronary artery disease; renal transplantation; human immunodeficiency virus (HIV) and acquired immune deficiency syndrome (AIDS); mental illness; and other (mainly invasive) procedures. The majority (94) examined coronary artery disease . . . "; p. 246: " . . . the quality of the research limits the ability to draw conclusions as to the existence of bias. Specialist services for coronary heart disease have been most extensively studied and it appears that gender differences in use do occur along the management pathway that leads to angiography . . ."; p. 247: " . . . the topics identified may themselves reflect gender bias in the research agenda. The vast majority of the studies are concerned with gender differences in coronary artery disease, the major cause of mortality in men. In contrast, few studies examined gender differences in mental health services, an area in which female patients predominate with respect to minor psychiatric morbidity . . ."
	N of articles	138	
	Years	1990–1999	
	Countries	US, UK, Canada, Israel, Finland, Germany, Spain	
	Study designs	O	

Govender and Penn-Kekana (220), 2007	**Health outcomes**	HC	p. 4: "Evidence demonstrating impact of gender on patient-provider interactions was relatively hard to come by and synthesis (sic). Apart from some studies documenting different treatment received by men and women most of the studies described are descriptive, or evaluations of relatively small scale interventions . . ."; p. 7: "sexual and reproductive health services are particularly impacted by gender bias and discrimination and . . . particular attention needs to be paid to how these services are provided within the health system"; p. 35: "Much of the gender discrimination that appears to take place is almost unconscious, reflecting the norms of the society in which both the health worker and the patient are based . . ."
	N of articles	11	
	Years	1983–2005	
	Countries	Global	
	Study designs	O	
LaResche (221), 2011	**Health outcomes**	HC	p. 1871: "Gender disparities in the amount of healthcare use for pain may be partially explained by the experience of higher-intensity pain in women. Pain intensity also seems to be a major factor influencing treatment, especially the prescription of medications for acute pain. However, clinicians' gender stereotypes, as well as the clinician's own gender, appear to influence diagnostic and treatment decisions for more persistent pain problems."
	N of articles	58	
	Years	1995–2010	
	Countries	US, Canada, Sweden	
	Study designs	O (55), E (3)	
McDonald (222), 2012	**Health outcomes**	P, M	p. 4: "Studies consistently demonstrate that targets of [sexual harassment] SH experience a range of significant negative psychological, health and job-related outcomes. Mental and physical health consequences range from irritation and anxiety to anger, powerlessness, humiliation, depression and post-traumatic stress disorder."
	N of articles	9	
	Years	1997–2009	
	Countries	US; Canada	
	Study designs	O	

(*continued*)

TABLE 3.4: Continued

Author & Year	Characteristics of Empirical Articles Reviewed on Discrimination and Health		Key Findings
4) Sexual orientation/identity (heterosexism; anti-lesbian, gay, bisexual, transgender [LGBT]) (n = 9)			
—see also "multiple types": Krieger (1), 3 studies; Pascoe and Richman (185), 13 studies; Goto et al. (187), 6 studies			
Williamson (223), 2000	**Health outcomes**	M	—includes only lesbian and gay; p. 103: "Research involving HIV and AIDS represents the largest body of knowledge for testing the relationship between conceptualizations and measures of internalized homophobia and illness. The research produced has had rather mixed results."
	N of articles	11	
	Years	1990–2009	
	Countries	US; Australia	
	Study designs	O	
Dean et al. (224), 2000	**Health outcomes**	P, M, HB, HC	—includes LGBT; p. 103: "Homophobia and heterosexism play a role in the inadequate assessment, treatment, and prevention of lesbian, gay, and bisexual health problems. LGB individuals suffer from discrimination in housing, employment, and basic civil rights . . . Research on the effects of stigma, violence, social attitudes, and gender bias on the lives of transgender individuals is even less available than for LGB populations. Preliminary reports and existing studies suggest that the problems above may be even more severe for transgender individuals."
	N of articles	9	
	Years	1989–2005	
	Countries	US	
	Study designs	O	
Meyer (225), 2003	**Health outcomes**	M, HB	—includes only LGB; p. 690: ". . . LGB individuals are exposed to excess stress due to their minority position and . . . this stress causes an excess in mental disorders."
	N of articles	14	
	Years	1994–2001	
	Countries	US	
	Study designs	O	

Szymanski et al. (226), 2008	**Health outcomes**	P, M, HB	—includes only LGB; p. 542: "the empirical literature shows relations between levels of IH [internalized homophobia] and sexual identity development, disclosure of sexual orientation to others, difficulties with the coming-out process, self-esteem, depression, psychological distress, social support, psychosocial distress, physical health, intimate and sexual relationship quality, adherence to traditional gender role and feminist attitudes, traditional religious beliefs, and perceived career barriers (among lesbians). Mixed findings were also reported, especially with regard to substance use in LGB individuals and risky sexual behavior in men who have sex with men."
	N of articles	42	
	Years	1986–2006	
	Countries	US, Canada, Australia, England, Scotland, Finland	
	Study designs	O	
DeSantis (227), 2009	**Health outcomes**	HIV/AIDS	—includes only transgender; p. 366: "Many MTF [male-to-female] transgender persons have reported that they do not seek health care because of previous experiences of discrimination or blatant verbal abuse by HCPs [health care providers]…" p. 368: "About 40% of MTF transgender persons have experienced significant stigma and discrimination…[which] is reported to be more intense and psychologically damaging than negative reactions that are experienced by gay, lesbian, and bisexual people…In addition to the stigma and discrimination from HCPs as previously discussed, transgender women experience a significant amount of employment-related discrimination…"
	N of articles	8	
	Years	1998–2008	
	Countries	US	
	Study designs	O	
Newcomb and Mustanski (228), 2010	**Health outcomes**	M	—includes only lesbian and gay; p. 1019: "Thirty-one studies were meta-analyzed for the relationship between [internalized homophobia] IH and mental health (N=5831), revealing a small to moderate overall effect size for the relationship between the two variables….The relationship between IH and depressive symptomatology was stronger than the relationship between IH and symptoms of anxiety."
	N of articles	31 (meta-analysis)	
	Years	1987–2008	
	Countries	US, Canada	
	Study designs	O	

(*continued*)

TABLE 3.4: Continued

Author & Year	Characteristics of Empirical Articles Reviewed on Discrimination and Health		Key Findings
5) Disability (physical and/or mental) (n = 4)			
—see also "multiple types": Krieger (1), 1 study			
Burns (229), 2009	**Health outcomes**	M, HC	—focused on mentally disabled; pp. 23–24: "Both real and perceived prejudice against the mentally disabled within the health sector is a potent barrier to accessing care . . . the mentally disabled receive unequal treatment for co-morbid physical disorders in comparison to their mentally-well counterparts."
	N of articles	6	
	Years	2002–2008	
	Countries	US, S Africa, Brazil, Argentina, Canada	
	Study designs	O	
Thornicroft et al. (230), 2007	**Health outcomes**	M, HC	—focus on mentally ill; p. 113: "This paper discusses factors associated with low rates of help-seeking and poorer quality of physical healthcare among people with mental illnesses"; p. 117: "A series of focus groups in England asked service users about their experiences of stigma and about who should receive targeted educational sessions to reduce discrimination. The group most often mentioned (by about two-thirds of service users) was family doctors, closely followed by school children, employers and police"; p. 118: "There is strong evidence that people with a diagnosis of a mental illness . . . receive inferior care for diabetes and heart attacks"; p. 119: "Surprisingly little has been written about why people stop attending appointments, from the point of view of service users themselves, either to understand what affects their satisfaction with care, or to find out how stigma and discrimination play a role in these treatment decisions, or to appreciate the mixed feelings many people with mental illness have about their harmful healers."
	N of articles	12	
	Years	1997–2005	
	Countries	US, UK	
	Study designs	O	

Emerson et al. (231), 2009	**Health outcomes**	P, M, HB, HC	—includes physically and mentally disabled; pp. 8–9: "Direct effects of discrimination on health result from the existence of systemic disability discrimination in the operation of health care systems, the direct consequence of which is to impede disabled people's access to appropriate health and social care. Indirect effects of discrimination on health arise through two interconnected pathways. First, discriminatory systems and practices contribute to the social exclusion of disabled people. As a result, disabled people are more likely than their peers to be exposed to living conditions (poverty, unemployment, social adversity, low control, low status, poor housing) associated with poor health. Second, the direct and indirect experience of disablism and disability discrimination may be expected to have a negative impact on the person's mental and physical health. While no direct evidence is available to support this contention, equivalent processes (the experience of racism and racial discrimination) have been identified as central to understanding ethnic inequalities in health."
	N of articles	5	
	Years	1999–2008	
	Countries	England	
	Study designs	O	
6) Age (n = 3)			
Ory et al. (232), 2003	**Health outcomes**	P, M, HB, HC	p. 164: "Ageist stereotypes are pervasive in U.S. society and harmful to older adults' psychological well-being, physical and cognitive functioning, and survival"; p. 166: "Ageism in medical care is manifested in doctors' tendency to give less aggressive treatments based on age characteristics alone, regardless of how the older person would actually fare with the withheld treatments or regimens; p. 166: "older people . . . exposed to negative stereotypes had reduced memory performance, self-efficacy, and the will to live" and "heightened cardiovascular response to stress."
	N of articles	7	
	Years	1996–2003	
	Countries	US	
	Study designs	O, E	

(*continued*)

TABLE 3.4: Continued

Author & Year	Characteristics of Empirical Articles Reviewed on Discrimination and Health		Key Findings
Bugental and Hehman (233), 2007	**Health outcomes**	P, M, HB, HC	p. 173: "...biased responses of others foster older adults' stress-related emotional and hormonal responses—responses that may effectively reduce their social and cognitive competence"; p. 180: "employees who undergo unwanted retirement show higher rates of morbidity and mortality in the years immediately following retirement, in comparison with those who undergo forced retirement but are then rehired"; p. 193: for elder abuse in the US and UK, "the highest prevalence rates were reported for neglect (#1) and financial exploitation (#2). Psychological and physical abuse occurred at slightly lower levels (at equivalent rates) across the two countries. The lowest rates were found for sexual abuse."
	N of articles	16	
	Years	1990–2007	
	Countries	US, UK	
	Study designs	O, E	
Meisner (234), 2012	**Health outcomes**	HC	p. 69: "From these published papers, predominantly representing an American system, it appears that physicians' attitudes toward aging are complex, dependent on contextual factors, and may be negative in content...The little research on aging attitude and aging knowledge generates more questions than answers"; p. 70: "Most of these themes that emerged from the literature are rooted within larger systemic issues that relate to the health care system itself...attitudes toward a decentralized health care system, the factors influencing financial reimbursement, and paper work may be stronger predictors of geriatric care attitudes than attitudes toward aging."
	N of articles	25	
	Years	1971–2009	
	Countries	US, Canada, Singapore	
	Study designs	O	

Abbreviations: (a) Health outcomes: P = physical health, M = mental health, HB = health behavior, HC = health care; (b) study design: O = observational; E = experimental

Search protocol:

1) data bases searched: Web of Science (2013) and PubMed (2013); the search was for review articles published between January 1, 1900 and June 1, 2013.

2) articles needed to be identified by the data base as a "review" and were selected for inclusion if they provided a review of empirical investigations that explicitly used measures of discrimination (structural or at individual level) to analyze health outcomes or else evaluated instruments used to measure discrimination in health research; conceptual reviews were not included, and the N of studies listed as reviewed refer only to the empirical studies regarding associations between exposure to discrimination and health (as opposed to background articles, or studies of discrimination with no health outcomes, etc.). Articles were not included if they focused only on stigma.

3) initial search: conducted on May 31, 2013, and restricted to articles published from January 1, 2000, through May 31, 2013.
core search terms = "discrimination OR bias OR prejudice" AND "health OR disease OR morbidity OR mortality"
and additional phrases added to the core were: (a) "racism OR racial OR race OR ethnic OR ethnicity OR indigenous OR native"; (b) "gender OR sexism"; (c) "anti-gay OR homophobia OR homophobic OR transgender OR transsexual OR LGBT OR (sexual AND minority) OR gay OR lesbian OR bisexual OR queer OR two-spirit"; (d) "class OR socioeconomic OR poverty"; (e) "disability OR disabled"; (f) "immigrant OR xenophobia OR nativist"; (g) "religion or religious"; (h) "age OR ageism"

4) based on review of initial results, added relevant articles cited in the bibliographies, and also conducted a second set of searches on June 5, 2013, and June 6, 2013, expanding dates of articles to January 1, 1900, through June 1, 2013
additional search terms = (a) "prejudice OR racism OR sexism OR homophobia OR heterosexism OR ableism OR ageism" AND "health OR disease OR mortality OR morbidity"; (b) "prejudice OR racism OR sexism" AND "Latino OR black OR (African AND American) OR Asian OR Pacific OR (American AND Indian) OR (Native AND American) OR Indigenous OR minority" AND "health OR disease OR mortality OR morbidity"; (c) "prejudice" AND "gender OR sex OR disabled OR impaired" AND "health OR disease OR mortality OR morbidity"; (d) "'residential OR occupational' AND segregation" OR "structural AND violence" and "health OR disease OR disability OR morbidity OR mortality"; (e) "environmental OR environment" AND "racism OR justice OR injustice" AND "health OR disease OR disability OR morbidity OR mortality"; (f) "incarceration OR prison OR jail OR (criminal AND justice)" AND "racism OR discrimination" AND "health OR disease OR disability OR morbidity OR mortality"; (g) "gender AND equity" AND "discrimination OR prejudice OR bias" AND "health OR disease OR morbidity OR mortality OR disability".

economic resources relevant to maintaining health. None of the reviews focused exclusively on anti-immigrant discrimination (although, as noted below, this type of discrimination was addressed in diverse articles concerned with racial discrimination), nor did any focus directly on religious discrimination. Finally, only 5 review articles encompassed multiple types of discrimination, all of which considered discrimination based on race/ethnicity and gender, 3 of which also included discrimination based on sexual orientation, and 2 of which addressed discrimination based on disability and age.

Four key features of contemporary empirical research on discrimination and health stand out (Table 3.4):

1. The vast majority of review articles and studies are focused on interpersonal discrimination, with the majority of research still focused on racial/ethnic discrimination:
   a. in the case of etiology, these studies primarily conceptualize discrimination as a stressor (i.e., type of social trauma), as measured directly using self-report instruments in observational studies, and
   b. in the case of medical treatment, they primarily use indirect methods, with discrimination inferred if observed group differences persist even after accounting for major known factors that potentially could influence treatment decisions (e.g., age, type of symptoms, stage of disease, comorbidity, etc.).
2. In parallel, the growing use of experimental methods chiefly is, for etiologic studies, mainly focused on psychoneurophysiological responses to adverse stimuli involving discriminatory treatment, and for health care studies, on medical decision-making in relation to use of vignettes and also unconscious bias.
3. Most review articles and the studies they include, whether observational or experimental, focus on only one type of discrimination; a growing number, however, employ self-report data on "unfair treatment" without specific attribution to any particular type of discrimination.
4. Only a handful of review articles focus on institutional or structural discrimination, and do so chiefly in relation to residential segregation and environmental racism.

Notably, the conclusions offered by the current review articles (Table 3.4) are in keeping with those of the first wave of investigations (1). Specifically:

1. The most robust etiologic findings pertain to positive associations between discrimination and psychological distress;
2. Growing evidence links exposure to discrimination to increased likelihood of adverse health behaviors (e.g., adverse use of psychoactive substances such as tobacco, alcohol, and other drugs, also unsafe sex);
3. Evidence for associations between discrimination and somatic health remains inconsistent and weak, whether for the still dominant work focused on cardiovascular outcomes (for which the evidence is stronger for cardiovascular reactivity than it is for hypertension), for the smaller number of new studies analyzing immunological and hormonal biomarkers of stress response, or for the handful of studies focused on obesity and on other noncommunicable and also infectious disease outcomes; and

4. Indirect and increasingly experimental evidence supports the hypothesis that decisions of health care providers can be adversely affected by bias (unconscious as well as conscious).

Does this accounting, however, fully capture the toll of discrimination on health and well-being? As informed by an ecosocial analysis, the likely answer is: no—for reasons which I will now elaborate.

ADVANCING THE WORK ON DISCRIMINATION AND HEALTH INEQUITIES

METHODOLOGICAL CHALLENGES FOR ANALYZING STRUCTURAL ASPECTS OF DISCRIMINATION FOR HEALTH RESEARCH

One striking and disturbing finding revealed by Table 3.4 is the paucity of research on structural or institutional discrimination as a determinant of health inequities (3, 8, 82). As previously discussed, however, discrimination is not an individual matter, even as one manifestation is through interpersonal encounters. Discrimination instead is at core a historically entrenched cross-generational societal phenomenon, one that creates and preserves privilege for dominant groups at the expense of subordinated groups. After all, if discrimination served no function, it would presumably be simple to eliminate.

State-sanctioned discrimination, past and present, is of particular concern (8). Consider the example of racism in the United States. Not surprisingly, because the rising pressure of the civil rights movement finally forced the US federal government to abolish legal (i.e., de jure) racial discrimination in the mid-1960s (98, 99), most contemporary US research on institutional racism and health (Table 3.4) primarily focuses on present day de facto discriminatory policies and practices, chiefly in relation to *(1)* residential, educational, and (to a lesser extent) occupational segregation and *(2)* environmental racism, as shaped by broader issues of political economy, political disempowerment, and poverty (100).

RECKONING WITH CURRENTLY LEGAL DISCRIMINATION: LIFETIME CONSEQUENCES

An important gap in current research, however, rendered visible by ecosocial theory's emphasis on accountability and agency, concerns the racialized health consequences of contemporary legal discrimination. Underscoring this point is nascent work on the myriad consequences of the legally color-blind, albeit racially motivated, US War on Drugs and its role in producing or exacerbating health-debilitating racial/ethnic inequalities (8, 101–106). As Alexander explains (107),

> President Ronald Reagan officially declared the current drug war in 1982, when drug crime was declining, not rising. From the outset, the war had little to do with drug crime and nearly

> everything to do with racial politics. The drug war was part of a grand and highly successful Republican Party strategy of using racially coded political appeals on issues of crime and welfare to attract poor and working class white voters who were resentful of, and threatened by, desegregation, busing, and affirmative action. In the words of H.R. Haldeman, President Richard Nixon's White House Chief of Staff: "[T]he whole problem is really the blacks. The key is to devise a system that recognizes this while not appearing to."

Consequently, despite substantial evidence that rates of illicit drug use are similar across all US racial/ethnic groups (101, 108: Table 58), research repeatedly has shown that African Americans are especially much more likely than white Americans to be arrested, convicted, and sentenced for use of drugs (101, 103, 105, 109). For example, a national study published in 2013 found that despite equal marijuana use by black and white Americans, black Americans were 3.7 times more likely, on average, to be arrested on charges of marijuana possession, with this excess risk ranging from a "low" of 2.5 times higher in some states (e.g., Colorado, Oregon) to over 5 times higher in other locales (e.g., Illinois, Iowa, District of Columbia) (109). Racial inequalities in imprisonment rates in the United States consequently exhibit not only period but also cohort effects, whereby the lifetime cumulative risk of imprisonment among US men age 30–54 born between 1945 and 1949 equaled 1.4% for white men versus 10.6% for black men, but among those born between 1965–1969, these values respectively equaled 2.9% and 20.5%, with lifetime risk of imprisonment among black men age 30–34 without a college degree in 1999 equal to 30.2% (as compared to 12.0% in 1979) (104). As stated in one recent review, as of 1999, "among black male high school dropouts, the risk of imprisonment had increased to 60%, establishing incarceration as a normal stopping point on the rout to midlife" (104: p. 164).

The health impact of racial discrimination in drug arrests and other arrests linked to institutional and structural discrimination (e.g., racial profiling for "stop-and-frisk" policing (2, 32), moreover, does not start and end with adverse exposure to health-damaging conditions in prison (8, 101, 106, 110). Following release, ex-felons are subject to legal discrimination in many US states, whereby they are not only denied the right to vote and serve on juries but also confront legal prohibitions limiting access to such well-known determinants of health as employment, housing, education, and public benefits (101, 103, 105, 110). The exclusion of prisoners from most health studies in turn leads to a type of selection bias that would result in civilian-based studies (including most national surveys) underestimating the extent of—and contribution of racial discrimination to—racial/ethnic health inequities (8, 103). Far from unique to the United States, these concerns are of global significance, given links in many countries between racism, risk of imprisonment, and health inequities (111).

Other examples of active legal discrimination in the United States involve sexual orientation. At issue is still-legal discrimination in employment and housing in states whose civil rights laws do not explicitly include protection on the basis of sexual orientation (34, 35), as well as now highly contested US state laws prohibiting gay marriage (as of June 2013, gay marriage was banned in 36 states, legal in 12 states plus DC, and neither authorized nor prohibited in 2 [22]), with the long-standing restriction of federal tax benefits to only heterosexual married couples (34) only overturned in late June 2013 (112). A series of studies, for example, has found that rates of psychiatric disorders among LGB persons, controlling for other relevant covariates: (1) were higher among LGB persons who resided in states that did not versus did extend protection against hate

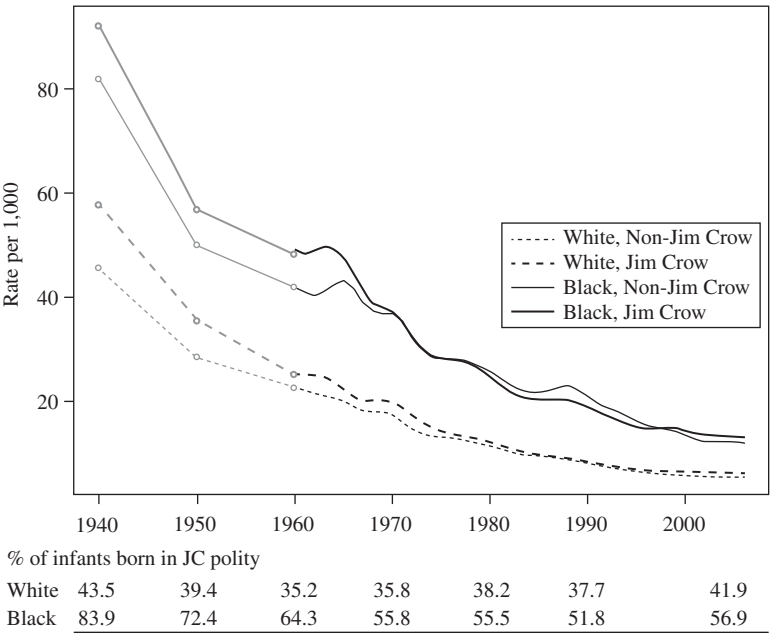
crimes and employment discrimination based on sexual orientation (113), and (2) increased in states that instituted bans on gay marriage (114); by implication, state protection of rights reduces health inequities. Active contests in many other countries regarding LGBT rights (20, 34, 35), including the right to gay marriage (as of mid-June 2013, now legal in 14 countries) (21), again suggests these US findings are likely relevant in other country contexts.

RECKONING WITH PRIOR LEGAL DISCRIMINATION: THE LONG REACH OF HISTORY, WITHIN AND ACROSS GENERATIONS

Nor is history dead within us. As ecosocial theory clarifies, measuring only contemporary exposure is likely to dilute estimates of the impact of discrimination on health (4, 8, 57). Age, period, and cohort effects all matter. A case in point is the mid-1960s abolition of US Jim Crow laws—that is, laws enacted in the late 19th and early 20th centuries that upheld white supremacy and sanctioned legal racial discrimination (predominantly against black Americans, but also affecting American Indians, Latinos, and Asian Americans) in voting, education, employment, health care, housing, the legal system, and use of public facilities, spaces, services, and transportation (17, 19, 98). In light of mounting evidence of the importance of early life conditions and cumulative disadvantage for both adult health and transgenerational transmission of risk (82, 83, 103, 115), a reasonable hypothesis is that Jim Crow, as well as its abolition, had both immediate and enduring health consequences. In 2013, all US-born persons aged 49 and older were born, and those aged 69 and above (i.e., the age group in which the bulk of mortality occurs) had already come of age and lived the first 20 years of their lives, and perhaps had their first child, when Jim Crow was legal in 21 out of 50 states plus the District of Columbia, with de facto discrimination in the remaining 29 states (8).

Yet, to date, scant research has investigated the impact of Jim Crow laws—or their abolition—for present-day racial/ethnic health disparities (8). Results of the literal handful of 5 studies of the topic have nevertheless all provided provocative evidence that the abolition of Jim Crow had a beneficial impact on black health, especially for infant mortality, and also a reduction, for some health outcomes, in black versus white health inequities, likely through a combination of improving working and living conditions plus the immediate impact of desegregation of hospital facilities (116–120). Even so, as indicated by Figure 3.4, although abolition of Jim Crow led to a singular convergence of black infant death rates comparing states that did versus did not have Jim Crow laws, its abolition was insufficient to eliminate the entrenched two-fold excess risk that continues to this day (120). A parallel argument regarding the continuing relevance of past as well as current injustice, including land expropriation, appears in the literature on Indigenous people's health, concerning the ongoing somatic and mental health consequences of historical trauma (121–124), a concept itself first developed to understand health outcomes among children of Holocaust survivors (121, 125).

Conversely, as implied by the examples of the abolition of Jim Crow and also the passage of laws preventing anti-LGBT discrimination, it likewise is essential that research on discrimination and health investigate the health impact of societal actions to end and also redress discrimination. As I noted in my original review (1), research on discrimination and health would likely benefit from engaging with work in the fast developing field of health and human rights (9, 10). The international human rights instruments listed in Table 3.2, for example, provide important benchmarks for assessing how enforcement (not just violation) of these internationally stipulated rights affects



% of infants born in JC polity							
White	43.5	39.4	35.2	35.8	38.2	37.7	41.9
Black	83.9	72.4	64.3	55.8	55.5	51.8	56.9

FIGURE 3.4: Infant death rates for US black and white infants in the Jim Crow (JC) and non–Jim Crow polities, 1940–1960: rates per 1000 (rolling 3-year average).

Note: Gray lines based on pretabulated US decennial mortality rates; black lines based on annual mortality data.

Source: Krieger et al. (120).

population health, both on average and in relation to the magnitude of health inequities. From a policy perspective, this could be particularly useful, since popular movements and professional organizations can hold governments, and sometimes even nonstate actors, accountable for stipulations in these human rights instruments (9, 10, 14). The troubling lack of any discussion of human rights in any of the post-2000 review articles included in Table 3.4, however, is not altogether surprising, given the previously noted neglect of epidemiologic research on political determinants of health (3). Related, only a handful of the review articles (primarily those regarding environmental justice and Indigenous health) mentioned community organizing and social movements for change. The gaps are many, and there is much work to do.

MEASUREMENT OF INDIVIDUALS' EXPOSURE TO DISCRIMINATION: EXPLICIT AND IMPLICIT

Despite the glaring need for rigorous research on structural discrimination and health, addressing concerns regarding the plausibility of presumed biological pathways requires credible investigation on individuals' exposures and their embodiment (8). A starting point is to remember that individual-level data are more than simply individual—precisely because no one is an "individual" one day and a member of a "population" another. Each person is both, simultaneously (58). At issue are not only people's individual experiences, both material and psychological, but also their

reference points for evaluating them—and for acting to alter future risk. Identification of discrimination at the individual level requires group-level knowledge—whether knowledge about group mores for what constitutes dignified treatment versus the denial of dignity (12, 126) or, in the case of discrimination regarding wages, occupational hazards, and medical referrals, knowledge about what others have experienced (1, 12). As the review articles make clear, methodological challenges remain for measuring individuals' exposure to discrimination.

INDIVIDUAL EXPLICIT (SELF-REPORT) DATA: DOMAINS AND ATTRIBUTION

Domains Matter

Two distinct types of instruments appear in the literature on discrimination and health (Table 3.4): *(1)* explicit measures of exposure to diverse domains, and *(2)* measures that emphasize psychosocial aspects of interpersonal interactions with less or no information about where the interactions occurred (8). For example, in relation to racial discrimination, widely used psychometrically validated examples of the former are *(1)* the Experiences of Discrimination (EOD) measure, which asks about discrimination in nine domains ("at school"; "getting hired or getting a job"; "at work"; "getting housing"; "getting medical care"; "getting service in a store or restaurant"; "getting credit, bank loans, or a mortgage"; "on the street or in a public setting"; and "from the police or in the courts") and also about people's responses to unfair treatment (127, 128), and *(2)* the major discrimination component of the Everyday Discrimination Scale (EDS), which specifies six domains (work, police, education, housing, bank, receipt of services) among its nine items (129). Also part of the EDS is a measure of "day-to-day unfair treatment" (129), which focuses on various types of unfair treatment, with the two domains mentioned (in the 10 items) pertaining to stores and restaurants.

Recently, a growing number of researchers, however, have begun using this latter EDS measure, regarding day-to-day unfair treatment, on its own, rather than in conjunction with the domain-oriented major discrimination EDS subscale (130–134). Yet, from both a data quality and also a prevention and policy perspective, asking about the multiple domains in which discrimination occurs is critical as a key complement to, not replacement for, questions that focus on psychosocial aspects of the exposure (8). In part, this is because specification of domains is important for cognitively grounding the question and response (12) and because critical theoretically informed review of such lists can reveal gaps and hence potential new domains for inclusion (e.g., racial discrimination in cyberspace [82]). Beyond any psychometric considerations, however, are key points pertaining to agency and accountability: the occurrence of discrimination in diverse domains, such as discrimination at work, in housing, in education, and in health care, is legally actionable (1, 12, 135), and knowing where discrimination occurs, as opposed to treating it only as a free-floating psychosocial stressor, is relevant to ending it (8).

Attribution Matters

Current instruments to assess exposure to discrimination also differ in how they ask their questions. The primary two main approaches respectively: *(1)* ask explicitly about discrimination

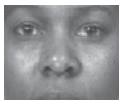
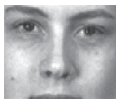
in the stem of the question (as in the Experiences of Discrimination measure [127, 128]), or *(2)* ask first about unfair treatment, and if any is reported, follow up with a question about attribution, for example, to race/ethnicity or something else (as in the Everyday Discrimination Scale measure [129]). As has been noted for over a decade (1, 136–138), these approaches are not equivalent.

Attesting to differences in these two approaches, in the case of racial discrimination, new empirical data from the 2007 California Health Interview Study unambiguously demonstrate—by employing identically phrased questions and a split-sample design—that self-reports of unfair treatment (without any attribution) are much higher—and demonstrate far less racial/ethnic variation—than self-reports of unfair treatment attributed to race/ethnicity and self-reports in response to a 1-stage question that asks directly about racial discrimination (137, 138). By implication, unfair treatment (without attribution) would contribute less to explaining racial/ethnic health disparities and also underestimate the health impact of racial discrimination. Supporting this inference, recent analyses from the Jackson Heart study (with 5,301 African American participants) found that whereas risk of hypertension was associated with higher lifetime self-reports of discrimination (whether attributed to race/ethnicity or not), it was only associated with the burden of discrimination (referring to appraisal of how stressful it was) when discrimination was attributed to race/ethnicity, and it was not associated with everyday discrimination (unattributed) (139). It is thus worrisome that empirical studies and review articles continue to treat findings arrived at through these two different methods as if they were directly comparable (Table 3.4; see also Lewis et al. [134]; Albert et al. [140]; Taylor et al. [141]). The larger issue raised by these findings is whether self-report data are adequate for measuring exposure to discrimination.

INDIVIDUAL IMPLICIT DATA

One of the newer approaches in the discrimination and health literature that seeks to minimize well-known cognitive problems affecting self-report data is the Implicit Association Test (IAT), a methodology initially developed to measure prejudice (41, 142–144). First used in health research to measure unconscious bias in health care providers and its effect on treatment decisions (145–147), the motivation for adapting the IAT for measuring exposure to discrimination (Figure 3.5) (148, 149) is the concern that the people most affected by discrimination may be least able or willing to say so, even as such experiences may nevertheless affect their health (1, 8, 127). Two lines of empirical evidence support this hypothesis.

First, the phenomenon psychologists refer to as the "person–group discrimination discrepancy" reveals that people typically report more discrimination for their group, on average, than for themselves personally—even though it is not possible for all individuals to experience, on average, less discrimination than their group (150, 151). Second, several studies observed a linear association between discrimination and health among more affluent persons, whereas among groups with fewer resources, risk was higher among respondents who reported no versus moderate discrimination, with the highest risk, however, occurring among respondents who reported high exposure (i.e., a J-shaped curve) (152–154). Together, these findings imply that self-reports of discrimination among exposed groups may underestimate exposure, especially among those

Discrimination:	Target concept categorization		Attribute categorization
Against Self (IAT-p)	Me My Mine	Them Their Theirs	Abuser Racist Bigot
Against Self (IAT-g)			Target Victim Oppressed

The IAT is a computer-based reaction-time methodology designed to capture phenomena that lie outside of the reaches of introspective access. The test contrasts the time it takes to make associations between two sets of items, e.g., "flowers" with the word "good," and "bugs" with the word "bad"–and then compares what happens when participants alternatively are asked to pair "flower" with "bad" and "bugs" with "good". A difference in average matching speed for opposite pairings determines the IAT score. Participants are typically aware that they are making these connections but unable to control them given the rapid response times and structure of the test. More than 500 studies have employed numerous versions of the IAT and have found the results to be robust, especially for phenomena that are subject to social desirability. Translated to the measurement of racial discrimination, as per the illustration above, we had two sets of target for the IAT. First, for discrimination against oneself, the measure – which we call the "IAT-p" (for person) – used the pronouns me, my, mine, them, their, and theirs. Second, for discrimination against one's group – which we call the "IAT-g" (for group) – we used photos of black and white persons. For both measures, the attribute categorization words were: abuser, racist, bigot, target, victim, and oppressed. Using these measures, we could ascertain the differences in strength of association for being a perpetrator versus target of discrimination.

FIGURE 3.5: Implicit Association Test (IAT) and use for measuring exposure to racial discrimination.
Sources: Carney et al. (148); Krieger (8); Krieger et al. (149).

with the least resources, even as this exposure can still adversely affect their health; one consequence would be underestimation of the impact of discrimination on health (1, 8, 57).

Tellingly, the first two studies to use the IAT to measure exposure to discrimination, both focused on racial discrimination (148, 149), have already shown that *(1)* the implicit measure does not detect the person–group discrimination discrepancy observed with the explicit measure, suggesting that this phenomenon reflects self-presentational bias, and *(2)* the correlation between implicit and explicit measures is small, implying that they capture different phenomena, with the low correlation on par with that reported in other social psychological research comparing implicit versus explicit measures of phenomena subject to self-presentational bias.

The second study also reported two notable health-related findings (149). First, the IAT and the Experiences of Discrimination responses were independently associated with risk of hypertension among black Americans. Second, in models comparing the black and white participants that controlled for age, gender, socioeconomic position (educational level of the respondent and both parents), body mass index, social desirability, and response to unfair treatment, black participants remained at significantly higher risk of being hypertensive (odds ratio [OR] = 1.4; 95% confidence interval [CI] = 1.0, 1.9). Their excess risk, however, was effectively eliminated and rendered statistically nonsignificant (OR = 1.1; 95% CI = 0.7, 1.7) by additionally adjusting for exposure to racial discrimination by using both the IAT and the Experiences of Discrimination measure.

These preliminary results thus point to the likely utility of health research on discrimination supplementing self-report data with IAT data (8).

EMBODYING EXPOSURE TO MULTIPLE TYPES OF DISCRIMINATION

Further underscoring the need for a more critical and integrated approach to investigating discrimination and health is the ecosocial inverse hazard law, modeled after Tudor Hart's famous inverse care law (155), and which posits that "the accumulation of health hazards tends to vary inversely with the power and resources of the populations affected" (156). At issue is the cumulative embodiment of multiple types of discrimination, deprivation, and other harmful exposures (8).

Although the review articles in Table 3.4 addressing multiple types of discrimination acknowledge the need for such integrated research, most empirical investigations continue either to focus on one type of discrimination at a time or else, as noted above, sometimes lump all types together under the common rubric of "unfair treatment." Among the earlier studies to question this assumption was a mid-1990s investigation that found that lesbian and gay African Americans reported higher rates of depressive distress than would be predicted based on summing risk for their race/ethnicity, gender, and sexual orientation (157). Also germane is new research on immigration and discrimination, which finds that recent US immigrants of color are the least likely to report having experienced racial discrimination, despite their greater likelihood in encountering discrimination based on language (138, 158–164). This finding should not be surprising because if, indeed, "race" is a social construct, it follows that people born and raised outside of the US have to learn how race is produced in the US and what US racial discrimination is like (158–160). This differential reporting, in conjunction with possibilities of a "healthy immigrant" effect (at least for the first generation) (165–167), points to the perils of ignoring nativity when assessing the impact of any kind of discrimination and health. This latter concern is of global relevance, in light of rising anti-immigrant discrimination in many countries across the world, variously construed in ethnic and religious terms (167, 168).

An empirical demonstration of why such an embodied approach is needed is analysis of data from the *United for Health* study, a cross-sectional investigation that recruited predominantly lower income women and men employees from diverse racial/ethnic groups, both US born and foreign born, from several workplaces in the Greater Boston area during 2003 and 2004 (169). Among members of this study, we documented high exposure to *(1)* socioeconomic deprivation, *(2)* occupational hazards (i.e., chemicals, dusts, fumes, and ergonomic strain), *(3)* social hazards (i.e., racial discrimination, workplace abuse, and sexual harassment at work), and *(4)* relationship hazards (i.e., intimate partner violence and unsafe sex) (170–173). Despite being union members, one-third of the study participants earned less than a living wage (equal to $10.54/hour at the time of the study) and 40% were below the US poverty line, while the black and Latino compared to white workers were nearly twice as likely to be impoverished (170).

Fully 85% of study members reported at least one high exposure to occupational hazards in the past year; nearly half (46%) reported three or more high exposures, and 17% reported five or more high exposures. Although some variation existed by race/ethnicity and gender, the majority of workers in each racial/ethnic-gender group were highly exposed (171, 172). Simultaneously, more

than 85% of the participants reported exposure to at least one of the three social hazards; exposure to all three reached 20% to 30% among the black workers, the most highly exposed group (170). Additionally, a substudy showed that among the black participants, immigrants reported less discrimination than their US counterparts, although this difference diminished with increasing time of the immigrants' residence in the US (164). For sexual harassment, an additional social category was relevant: sexuality. Specifically, lesbian, gay, bisexual, and transgender workers reported twice as much sexual harassment as did their heterosexual counterparts (170). Furthermore, within each racial/ethnic group, about one-third of the men reported having ever been a perpetrator of intimate partner violence, and about one-third of the women reported having been a target of such violence (170).

As exemplified by analyses of severe psychological distress, attaining an accurate picture of risk required considering all the social hazards together. Findings revealed that analyses that included data on only one type of hazard yielded estimates of risk biased by not taking the other types of hazards into account. Moreover, analyses including all three hazards demonstrated the especially high toll imposed by racial discrimination, independent of other exposures (173).

TOWARD A RIGOROUS SCIENCE OF RESEARCH ON DISCRIMINATION AND HEALTH INEQUITIES

In conclusion, as this chapter demonstrates, rigorous scientific study of discrimination and health inequities requires: *(1)* conceptual clarity about the exploitative and oppressive realities of adverse discrimination; *(2)* careful attention to domains, pathways, level, and spatiotemporal scale, in historical context; *(3)* structural-level measures; *(4)* individual-level measures, albeit without relying solely on self-report data or reducing discrimination to solely a psychosocial exposure; and *(5)* an embodied analytic approach, one attuned to biological expression of historically contingent and dynamic societal conditions and also to how discrimination can adversely affect the production of scientific knowledge itself.

Stated simply, the epidemiology of the health consequences of discrimination is, at heart, the investigation of intimate connections between our social and biological existence. It is about how truths of our body and body politic engage and enmesh, thereby producing population patterns of health, disease, and well-being. To research how discrimination harms health, we accordingly must draw on not only a nuanced understanding of the likely biological pathways of embodying discrimination, from conception to death, but also a finely tuned historical, social, and political sensibility, situating both the people we study and ourselves in the larger context of our times. Frank appraisal of diverse types of discrimination operative in any given country context are thus required, with research needed not only to deepen understanding of the health impact of types of discrimination already the subject of active research (e.g., racial discrimination) but also types of discrimination for which much less research exists (e.g., in relation to gender, sexuality, disability, age, social class, immigrant status, and religion), both singly and combined.

The stakes for getting our science right are high—both scientifically and practically. As this review of the state-of-the-field clarifies, extant research is likely to yield conservative, not inflated, estimates of the impact of discrimination on health. Partly this is because of the emphasis on discrimination as an interpersonal psychosocial stressor, with the conservative bias magnified by reliance primarily on self-report exposure data, including exposure measures that refer only to "unfair treatment" overall, without specifying type or domain of discrimination. Concomitantly, research on the impact of structural discrimination—and efforts to end it—is sorely lacking, limiting understanding of the toll of discrimination on people's health, its contribution to social inequalities in health, and how it can be altered. Although data by themselves cannot rectify health inequities, the absence of data demonstrating harm nevertheless is itself harmful (1, 4)—as underscored by the time-worn adage "no data, no problem" (174). Our responsibility, as public health researchers, is to use the best science possible—conceptually and methodologically—to build public clarity about the extent and health consequences of discrimination and the life-affirming value of eradicating it, as one necessary contribution to the society-wide task of shifting the balance from embodying inequity to embodying equity.

ACKNOWLEDGMENTS

This chapter draws on work from three prior publications (1, 8, 58) and thinking spurred by empirical research supported by several of my NIH-funded studies (1 R01 AG027122-01; R03 CA137666-01A1; 3R01AG027122-03S1; 1R21HD060828-01A2; 1 R21 CA168470-01; 3R21HD060828-02S1) and also a Robert Wood Johnson Society and Health seed grant.

REFERENCES

1. Krieger N. Embodying inequality: a review of concepts, measures, and methods for studying health consequences of discrimination. Int J Health Services. 1999;29:295–352. Republished and slightly updated as: Krieger, N. Discrimination and health. In: Berkman L, Kawachi I, eds. Social epidemiology. New York: Oxford University Press; 2000. 36–75.
2. Weiser B, Goldstein J. New York City asks court to vacate rulings on stop-and-frisk tactic. New York Times. 2013 November 10.
3. Beckfield J, Krieger N. Epi + demos + cracy: linking political systems and priorities to the magnitude of health inequities—evidence, gaps, and a research agenda. Epidemiol Rev. 2009;31(1):152–77.
4. Krieger N. Epidemiology and the people's health: theory and context. New York: Oxford University Press; 2011.
5. Longino HE. Studying human behavior: how scientists investigate aggression and sexuality. Chicago: University of Chicago Press; 2013.
6. Krieger N. Got theory? On the 21st c. CE rise of explicit use of epidemiologic theories of disease distribution: a review and ecosocial analysis. Curr Epidemiol Rep. 2013;1(1):1–12.
7. National Institutes of Health. Biennial Report of the Director, Fiscal Years 2008 & 2009. 2010 [June 16, 2013]. Available from: http://www.report.nih.gov/biennialreport0809/.

8. Krieger N. Methods for the scientific study of discrimination and health: an ecosocial approach. Am J Public Health. 2012;102(5):936–44.
9. Gruskin S, Mills EJ, Tarantola D. History, principles, and practice of health and human rights. Lancet. 2007;370(9585):449–55.
10. Grodin M, Tarantola D, Annas G, Gruskin S, editors. Health and human rights in a changing world. New York: Routledge; 2013.
11. World Health Organization Commission on the Social Determinants of Health (CSDH). Closing the gap in a generation: health equity through action on the social determinants of health. Final report of the Commission on Social Determinants of Health. Geneva: World Health Organization, 2008.
12. National Research Council. Measuring racial discrimination. Blank RM, Babady R, Citro CF, editors. Washington, DC: National Academies Press; 2004.
13. United Nations General Assembly. Universal declaration of human rights. Resolution 217A (III), Adopted and proclaimed December 10, 1948.
14. Tomasevski K. Women and human rights. London, UK: Zed Books; 1993.
15. Krieger N, Alegría M, Almeida-Filho N, Barbosa da Silva J, Barreto ML, Beckfield J, et al. Who, and what, causes health inequities? Reflections on emerging debates from an exploratory Latin American/North American workshop. J Epidemiol Community Health. 2010;64(9):747–9.
16. Oxford University Press. Oxford English Dictionary On-line. Available from: http://www.oed.com.ezp-prod1.hul.harvard.edu/.
17. Murray P. States' laws on race and color. Athens, GA: Women's Division of Christian Services; 1950.
18. Jaynes GD, Williams Jr. RM, editors. A common destiny: blacks and American society. Washington, DC: National Academy Press; 1989.
19. Anderson CE. Eyes off the prize: the United Nations and the African American struggle for human rights, 1944–1955. Cambridge, UK: Cambridge University Press; 2003.
20. Chauncey G. Why marriage? The history shaping today's debate over gay equality. New York: Basic Books; 2004.
21. Erlanger S. Hollande signs French gay marriage into law. New York Times. 2013 May 18.
22. Human Rights Campaign. Marriage Center [June 17, 2013]. Available from: http://www.hrc.org/campaigns/marriage-center.
23. Institute of Medicine. Unequal treatment: confronting racial and ethnic disparities in health care. Smedley BD, Stith AY, Nelson AR, editors. Washington, DC: National Academies Press; 2003.
24. Ruiz MT, Verbrugge LM. A two way view of gender bias in medicine. J Epidemiol Community Health. 1997;51(2):106–9.
25. De Vos P. Appendix I: introduction to South Africa's 1996 Bill of Rights. Netherlands Quarterly of Human Rights. 1997;15:225–52.
26. Sanders D, Chopra M. Key challenges to achieving health for all in an inequitable society: the case of South Africa. Am J Public Health. 2006;96(1):73–8.
27. Bond P. South African people power since the mid-1980s: two steps forward, one back. Third World Q. 2012;33(2):243–64.
28. Jary D, Jary J, editors. Collins dictionary of sociology. Glasgow, UK: HarperCollins; 1995.
29. Marshall G, editor. The concise Oxford dictionary of sociology. Oxford, UK: Oxford University Press; 1994.
30. Essed P. Understanding everyday racism: an interdisciplinary theory. London, UK: Sage; 1992.
31. Rothenberg P, editor. Race, class, and gender in the United States: an integrated study. 7th ed. New York: St. Martin's Press; 2007.

32. Reskin B. The race discrimination system. Annu Rev Sociol. 2012;38(1):17–35.
33. Sargent M. Age discrimination and diversity. Cambridge, UK: Cambridge University Press; 2011.
34. Badgett MVL, Frank J, editors. Sexual orientation discrimination: an international perspective. New York: Routledge; 2007.
35. Vaid U. Irresistible revolution: confronting race, class, and other assumptions of lesbian, gay, bisexual, and transgender politics. New York: Magnus; 2012.
36. Sennett R, Cobb J. The hidden injuries of class. New York: Knopf; 1972.
37. Crenshaw K. Mapping the margins: intersectionality, identity politics, and violence against women of color. Stanford Law Review. 1991;43(6):1241–99.
38. Garry A. Intersectionality, metaphors, and the multiplicity of gender. Hypatia. 2011;26(4):826–50.
39. Collins PH. Black feminist thought: knowledge, consciousness, and the politics of empowerment. London, UK: HarperCollins Academic Press; 1990.
40. Bobo L, Charles CZ, Krysan M, Simmons AD. The real records on racial attitudes. In: Marsden P, editor. Social trends in American life: findings from the General Social Survey since 1972. Princeton, NJ: Princeton University Press; 2012.
41. Banaji MR, Greenwald AG. Blind spot: hidden biases of good people. New York: Delacorte Press; 2013.
42. Pincus FL. Reversing discrimination: dismantling the myth. Boulder, CO: Lynne Rienner Publishers; 2003.
43. Herrnstein RJ, Murray C. The bell curve: intelligence and class structure in American life. New York: Free Press; 2010.
44. Thernstrom S, Thernstrom A. American in black and white: one nation, indivisible. New York: Simon & Schuster; 1997.
45. Satel SL. PC, MD: how political correctness is corrupting medicine. New York: Basic Books; 2000.
46. Jackman MR. The velvet glove: paternalism and conflict in gender, class, and race relations. Berkeley: University of California Press; 1994.
47. Jones RP, Cox D. Old alignment, emerging fault lines: religion in the 2010 election and beyond—findings from the 2010 post-election American Values Survey Washington, DC: Public Religion Institute; 2010 [June 17, 2013]. Available from: http://publicreligion.org/research/2010/11/old-alignments-emerging-fault-lines-religion-in-the-2010-election-and-beyond/.
48. Krieger N. Epidemiology and the web of causation: has anyone seen the spider? Soc Sci Med. 1994;39(7):887–903.
49. Krieger N. Theories for social epidemiology in the 21st century: an ecosocial perspective. Int J Epidemiol. 2001;30(4):668–77.
50. Krieger N. Embodiment: a conceptual glossary for epidemiology. J Epidemiol Community Health. 2005;59(5):350–5.
51. Krieger N. Shades of difference: theoretical underpinnings of the medical controversy on black/white differences in the United States, 1830–1870. Int J Health Services. 1987;17(2):259–78.
52. Ernst W, Harris B, editors. Race, science and medicine, 1700–1960. London, UK: Routledge; 1999.
53. Haller JS. Outcasts from evolution: Scientific attitudes of racial inferiority, 1859–1900. Urbana: University of Illinois Press; 1971.
54. Kevles DJ. In the name of eugenics: genetics and the uses of human heredity. New York: Knopf; 1985.
55. Haraway DJ. Primate visions: Gender, race, and nature in the world of modern science. New York: Routledge; 1989.
56. Ziman J. Real science: what it is, and what it means. Cambridge, UK: Cambridge University Press; 2000.
57. Krieger N. The science and epidemiology of racism and health: racial/ethnic categories, biological expressions of racism, and the embodiment of inequality—an ecosocial perspective. In: Whitmarsh I, Jones DS,

editors. What's the use of race? Genetics and difference in forensics, medicine, and scientific research. Cambridge, MA: MIT Press; 2010. p. 225–55.

58. Krieger N. Who and what is a "population"? Historical debates, current controversies, and implications for understanding "population health" and rectifying health inequities. Milbank Q. 2012;90(4):634–81.
59. Smith GD. Epidemiology, epigenetics and the "Gloomy Prospect": embracing randomness in population health research and practice. Int J Epidemiol. 2011;40(3):537–62.
60. Whitehead M. The concepts and principles of equity and health. Int J Health Services. 1992;22(3):429–45.
61. Braveman P. Health disparities and health equity: concepts and measurement. Annu Rev Public Health. 2006;27(1):167–94.
62. Krieger N. Defining and investigating social disparities in cancer: critical issues. Cancer Causes Control. 2005;16(1):5–14.
63. Galton F. Natural inheritance. London, UK: Macmillan; 1889.
64. Galton F. Eugenics: its definition, scope, and aims. Am J Sociol. 1904;10(1):1–25.
65. Stigler SM. Regression towards the mean, historically considered. Stat Methods Med Res. 1997;6(2):103–14.
66. Limpert E, Stahel WA, Abbt M. Log-normal distributions across the sciences: keys and clues. Bioscience. 2001;51(5):341–52.
67. Smith JM. On the fourteenth query of Thomas Jefferson's notes on Virginia. Anglo-African Magazine. 1859;1:225–38.
68. Reyburn R. Remarks concerning some of the diseases prevailing among the freedpeople in the District of Columbia (bureau refugees, freedmen, and abandoned lands). Am J Med Sci. 1866;51(102):364–9.
69. Dubois WEB. The health and physique of the Negro American. Atlanta, GA: Atlanta University Press; 1906.
70. Tibbitts C. The socio-economic background of negro health status. J Negro Educ. 1937;6(3):413–28.
71. Krieger N, Rowley D, Hermann AA, Avery B, Phillips MT. Racism, sexism, and social class: implications for studies of health, disease, and well-being. Am J Prev Med. 1993;9 (Suppl 6):82–122.
72. Williams DR, Collins C. US socioeconomic and racial differences in health: patterns and explanations. Annu Rev Sociol. 1995;21:349–86.
73. Williams DR, Mohammed SA, Leavell J, Collins C. Race, socioeconomic status, and health: Complexities, ongoing challenges, and research opportunities. Ann N Y Acad Sci. 2010;1186(1):69–101.
74. Shavers VL, Shavers BS. Racism and health inequity among Americans. J Natl Med Assoc. 2006;98:386–96.
75. Fix M, Struyk R. Clear and convincing evidence: measurement of discrimination in America. Washington, DC: Urban Institute Press; 1993.
76. Oliver ML, Shapiro TM. Black wealth/white wealth: 10th anniversary edition. New York: Routledge; 2006.
77. Gravlee CC. How race becomes biology: embodiment of social inequality. Am J Phys Anthropol. 2009;139(1):47–57.
78. Schuman H, Steehm C, Bobo L. Racial attitudes in America: trends and interpretations. Cambridge, MA: Harvard University Press; 1985.
79. Feagin JR, Sikes MP. Living with racism: the black middle class experience. Boston: Beacon Press; 1994.
80. Mays VM. Black women, women, stress, and perceived discrimination: the focused support group model as an intervention for stress reduction. Cult Divers Ment Health. 1995;1:53–65.
81. Bobo L, Zubrinsky CL, Johnson Jr. JH, Oliver ML. Work orientation, job discrimination, and ethnicity: a focus group perspective. Res Soc Work. 1995;5:45–85.
82. Gee GC, Ford CL. Structural racism and health inequities: old issues, new directions. Du Bois Rev. 2011;8(01):115–32.
83. Gee GC, Walsemann KM, Brondolo E. A life course perspective on how racism may be related to health inequities. Am J Public Health. 2012;102(5):967–74.

84. Armstead CA, Lawler KA, Gorden G, Cross J, Gibbons J. Relationship of racial stressors to blood pressure responses and anger expression in black college students. Health Psychol. 1989;8(5):541–56.
85. Jones DR, Harrell JP, Morris-Prather CE, Thomas J, Omowale N. Affective and physiological responses to racism: the roles of afrocentrism and mode of presentation. Ethn Dis. 1996;6:109–22.
86. Yankauer A, Jr. The relationship of fetal and infant mortality to residential segregation: an inquiry into social epidemiology. Am Sociol Rev. 1950;15(5):644–8.
87. LaVeist TA. The political empowerment and health status of African-Americans: mapping a new territory. Am J Sociol. 1992;97(4):1080–95.
88. LaVeist TA. Segregation, poverty, and empowerment: health consequences for African Americans. Milbank Q. 1993;71:41–64.
89. Wallace R, Wallace D. Socioeconomic determinants of health: community marginalisation and the diffusion of disease and disorder in the United States. BMJ. 1997;314(7090):1341.
90. Polednak AP. Segregation, poverty, and mortality in urban African Americans. New York: Oxford University Press; 1997.
91. Kennedy B, Kawachi I, Lochner K, Jones C, Prothrow-Stith D. (Dis)respect and black mortality. Ethn Dis. 1997;7(3):207.
92. Raudenbush SW. Hierarchical linear models: applications and data analysis methods. Thousands Oaks, CA: Sage Publications; 2002.
93. Gelman A. Data analysis using regression and multilevel/hierarchical models. New York: Cambridge University Press; 2007.
94. Sen G, Östlin P, the Women, Gender and Equity Knowledge Network. Unequal, unfair, ineffective and inefficient—gender inequity in health: why it exists and how we change it. Final report to the WHO Commission on the Social Determinants of Health, September. 2007 [June 17, 2013]. Available from: http://www.who.int/social_determinants/publications/womenandgender/en/index.html.
95. Connell R. Gender, health and theory: conceptualizing the issue, in local and world perspective. Soc Sci Med. 2012;74(11):1675–83.
96. Springer KW, Stellman JM, Jordan-Young RM. Beyond a catalogue of differences: a theoretical frame and good practice guidelines for researching sex/gender in human health. Soc Sci Med. 2012;74(11):1817–24.
97. Hawkes S, Buse K. Gender and global health: evidence, policy, and inconvenient truths. Lancet. 2013;381(9879):1783–7.
98. Fairclough A. Better day coming: blacks and equality, 1890–2000. New York: Viking; 2001.
99. Chafe WH, Gavins R, Korstad R, editors. Remembering Jim Crow: African Americans tell about life in the segregated South. New York: New Press; 2001.
100. Morello-Frosch RA. Discrimination and the political economy of environmental inequality. Environment and Planning C: Government and Policy. 2002;20(4):477–96.
101. Alexander M. The new Jim Crow: mass incarceration in the age of colorblindness. New York: The New Press; 2010.
102. Moore LD, Elkavich A. Who's using and who's doing time: incarceration, the war on drugs, and public health. Am J Public Health. 2008;98(9 Supp1):S176–S80.
103. London AS, Myers NA. Race, incarceration, and health: a life-course approach. Res Aging. 2006;28(3):409–22.
104. Pettit B, Western B. Mass imprisonment and the life course: race and class inequality in U.S. Incarceration. Am Sociol Rev. 2004;69(2):151–69.
105. Schnittker J, Massoglia M, Uggen C. Incarceration and the health of the African American community. Du Bois Rev. 2011;8(01):133–41.

106. Dumont DM, Brockmann B, Dickman S, Alexander N, Rich JD. Public health and the epidemic of incarceration. Annu Rev Public Health. 2012;33:325.

107. Alexander M. The new Jim Crow: how the war on drugs gave birth to a permanent American undercaste: Mother Jones; 2010 [June 17, 2013]. Available from: http://www.motherjones.com/politics/2010/03/new-jim-crow-war-on-drugs.

108. National Center for Health Statistics. Health, United States, 2012: With special feature on emergency care. Hyattsville, MD: NCHS, 2013.

109. American Civil Liberties Union. The war on marijuana in black and white: billions of dollars wasted on racially biased arrests New York: ACLU; 2013 [June 17, 2013]. Available from: https://www.aclu.org/files/assets/aclu-thewaronmarijuana-rel2.pdf.

110. Purtle J. Felon disenfranchisement in the United States: a health equity perspective. Am J Public Health. 2012;103(4):632–7.

111. Reynolds M. The war on drugs, prison building, and globalization: catalysts for the global incarceration of women. NWSA J. 2008;20(2):72–95.

112. Liptak A. Supreme Court bolsters gay marriage with two major rulings. New York Times. 2013 June 26.

113. Hatzenbuehler ML, Keyes KM, Hasin DS. State-level policies and psychiatric morbidity in lesbian, gay, and bisexual populations. Am J Public Health. 2009;99(12):2275–81.

114. Hatzenbuehler ML, McLaughlin KA, Keyes KM, Hasin DS. The impact of institutional discrimination on psychiatric disorders in lesbian, gay, and bisexual populations: a prospective study. Am J Public Health. 2010;100(3):452–9.

115. Kuh D, Ben-Shlomo Y, editors. A life course approach to chronic disease epidemiology: tracing the origins of ill-health from early to adult life. 2nd ed. Oxford: Oxford University Press; 2004.

116. Chay KY, Greenstone M. The convergence in black-white infant mortality rates during the 1960's. Am Econ Rev. 2000;90(2):326–32.

117. Almond D, Chay KY, Greenstone M. Civil rights, the war on poverty, and black-white convergence in infant mortality in the rural South and Mississippi. MIT Department of Economics: Working Paper No. 07-04; 2006 [June 17, 2013]. Available from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=961021.

118. Almond D, Chay KY. The long-run and intergenerational impact of poor infant health: evidence from cohorts born during the civil rights era. 2008 [June 17, 2013]. Available from: http://users.nber.org/~almond/chay_npc_paper.pdf.

119. Kaplan G, Ranjit N, Burgard S. Lifting gates, lengthening lives: did civil rights policies improve the health of African-American women in the 1960s and 1970s? In: Schoeni RF, House JS, Kaplan G, Pollack H, editors. Making Americans healthier: social and economic policy as health policy. New York: Russell Sage Foundation; 2008. p. 145–70.

120. Krieger N, Chen JT, Coull B, Waterman PD, Beckfield J. The unique impact of abolition of Jim Crow laws on reducing inequities in infant death rates and implications for choice of comparison groups in analyzing societal determinants of health. Am J Public Health. 2013;103(12):2234–44.

121. Brave Heart MY, DeBruyn LM. The American Indian holocaust: healing historical unresolved grief. Am Ind Alaska Native Mental Health Res. 1998;2:56–78.

122. Whitbeck L, Adams G, Hoyt D, Chen X. Conceptualizing and measuring historical trauma among American Indian people. Am J Community Psychol. 2004;33(3–4):119–30.

123. Carson B, Dunbar T, Chenhall RD, Bailie R. Social determinants of Indigenous health. Crows Nest, NSW, Australia: Allen & Unwin; 2007.

124. Walters KL, Mohammed SA, Evans-Campbell T, Beltrán RE, Chae DH, Duran B. Bodies don't just tell stories, they tell histories. Du Bois Rev. 2011;8(01):179–89.

125. Prince RM. Second generation effects of historical trauma. Psychoanal Rev. 1985;72(1):9–29.

126. Appiah KA. The honor code: how moral revolutions happen. New York: WW Norton & Company; 2010.

127. Krieger N. Racial and gender discrimination: risk factors for high blood pressure? Soc Sci Med. 1990;30(12):1273–81.

128. Krieger N, Smith K, Naishadham D, Hartman C, Barbeau EM. Experiences of discrimination: validity and reliability of a self-report measure for population health research on racism and health. Soc Sci Med. 2005;61(7):1576–96.

129. Williams DR, Yan Yu, Jackson JS, Anderson NB. Racial differences in physical and mental health: socio-economic status, stress and discrimination. J Health Psychol. 1997;2(3):335–51.

130. Schulz AJ, Gravlee CC, Williams DR, Israel BA, Mentz G, Rowe Z. Discrimination, symptoms of depression, and self-rated health among African American women in Detroit: results from a longitudinal analysis. Am J Public Health. 2006;96(7):1265–70.

131. Lewis TT, Everson-Rose SA, Powell LH, Matthews KA, Brown C, Karavolos K, et al. Chronic exposure to everyday discrimination and coronary artery calcification in African-American women: the SWAN heart study. Psychosom Med. 2006;68(3):362–8.

132. Gee GC, Spencer MS, Chen J, Takeuchi D. A nationwide study of discrimination and chronic health conditions among Asian Americans. Am J Public Health. 2007;97(7):1275–82.

133. Pérez DJ, Fortuna L, Alegría M. Prevalence and correlates of everyday discrimination among U.S. Latinos. J Community Psychol. 2008;36(4):421–33.

134. Lewis TT, Aiello AE, Leurgans S, Kelly J, Barnes LL. Self-reported experiences of everyday discrimination are associated with elevated C-reactive protein levels in older African-American adults. Brain Behav Immun. 2010;24(3):438–43.

135. US Department of Justice. Civil Rights Division. 2013 [June 17, 2013]. Available from: http://www.justice.gov/crt/.

136. Brown TN. Measuring self-perceived racial and ethnic discrimination in social surveys. Sociol Spectr. 2001;21(3):377–92.

137. Shariff-Marco S, Gee GC, Breen N, Willis G, Reeve BB, Grant D, et al. A mixed-methods approach to developing a self-reported racial/ethnic discrimination measure for use in multiethnic health surveys. Ethn Dis. 2009;19:447–53.

138. Shariff-Marco S, Breen N, Landrine H, Reeve BB, Krieger N, Gee GC, et al. Measuring everyday racial/ethnic discrimination in health surveys. Du Bois Rev. 2011;8(01):159–77.

139. Sims M, Diez-Roux AV, Dudley A, Gebreab S, Wyatt SB, Bruce MA, et al. Perceived discrimination and hypertension among African Americans in the Jackson Heart Study. Am J Public Health. 2012;102(S2):S258–S65.

140. Albert MA, Cozier Y, Ridker PM, Palmer JR, Glynn RJ, Rose L, et al. Perceptions of race/ethnic discrimination in relation to mortality among black women: results from the Black Women's Health Study. Arch Intern Med. 2010;170(10):896–904.

141. Taylor TR, Williams CD, Makambi KH, Mouton C, Harrell JP, Cozier Y, et al. Racial discrimination and breast cancer incidence in US black women: The Black Women's Health Study. Am J Epidemiol. 2007;166(1):46–54.

142. Greenwald AG, Nosek BA, Banaji MR. Understanding and using the Implicit Association Test: I. An improved scoring algorithm. J Pers Soc Psychol. 2003;85(2):197–216.

143. Greenwald AG, Poehlman TA, Uhlmann E, Banaji MR. Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. J Pers Soc Psychol. 2009;97(1):17–41.

144. Fazio RH, Olson MA. Implicit measures in social cognition research: their meaning and use. Annu Rev Psychol. 2003;54(1):297–327.

145. Green A, Carney D, Pallin D, Ngo L, Raymond K, Iezzoni L, et al. Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. J Gen Intern Med. 2007;22(9):1231–8.

146. van Ryn M, Burgess DJ, Dovidio JF, Phelan SM, Saha S, Malat J, et al. The impact of racism on clinician cognition, behavior, and clinical decision making. Du Bois Rev. 2011;8(01):199–218.

147. Shavers VL, Fagan P, Jones D, Klein WMP, Boyington J, Moten C, et al. The state of research on racial/ethnic discrimination in the receipt of health care. Am J Public Health. 2012;102(5):953–66.

148. Carney DR, Banaji MR, Krieger N. Implicit measures reveal evidence of personal discrimination. Self Identity. 2010;9(2):162–76.

149. Krieger N, Carney D, Lancaster K, Waterman PD, Kosheleva A, Banaji M. Combining explicit and implicit measures of racial discrimination in health research. Am J Public Health. 2010;100(8):1485–92.

150. Crosby F. The denial of personal discrimination. Am Behav Sci. 1984;27(3):371–86.

151. Taylor DM, Wright SC, Moghaddam FM, Lalonde RN. The personal/group discrimination discrepancy: perceiving my group, but not myself, to be a target for discrimination. Pers Soc Psychol Bull. 1990;16(2):254–62.

152. Krieger N, Sidney S. Racial discrimination and blood pressure: the CARDIA study of young black and white adults. Am J Public Health. 1996;86(10):1370–8.

153. Huebner DM, Davis MC. Perceived antigay discrimination and physical health outcomes. Health Psychol. 2007;26(5):627–34.

154. Chae DH, Lincoln KD, Adler NE, Syme SL. Do experiences of racial discrimination predict cardiovascular disease among African American men? The moderating role of internalized negative racial group attitudes. Soc Sci Med. 2010;71(6):1182–8.

155. Tudor Hart J. The inverse care law. Lancet. 1971;297(7696):405–12.

156. Krieger N, Chen JT, Waterman PD, Hartman C, Stoddard AM, Quinn MM, et al. The inverse hazard law: blood pressure, sexual harassment, racial discrimination, workplace abuse and occupational exposures in US low-income black, white and Latino workers. Soc Sci Med. 2008;67(12):1970–81.

157. Cochran SD, Mays VM. Depressive distress among homosexually active African American men and women. Am J Psychiatry. 1994;151(4):524–9.

158. Waters MC. Black identities: West Indian immigrant dreams and American realities. Cambridge, MA: Harvard University Press; 2001.

159. Hall SP, Carter RT. The relationship between racial identity, ethnic identity, and perceptions of racial discrimination in an Afro-Caribbean descent sample. J Black Psychol. 2006;32(2):155–75.

160. Deaux K, Bikmen N, Gilkes A, Ventuneac A, Joseph Y, Payne YA, et al. Becoming American: stereotype threat effects in Afro-Caribbean immigrant groups. Soc Psychol Q. 2007;70:384–404.

161. Araújo BY, Borrell LN. Understanding the link between discrimination, mental health outcomes, and life chances among Latinos. Hisp J Behav Sci. 2006;28(2):245–66.

162. Viruell-Fuentes EA. Beyond acculturation: immigration, discrimination, and health research among Mexicans in the United States. Soc Sci Med. 2007;65(7):1524–35.

163. Yoo HC, Gee GC, Takeuchi D. Discrimination and health among Asian American immigrants: disentangling racial from language discrimination. Soc Sci Med. 2009;68(4):726–32.

164. Krieger N, Kosheleva A, Waterman PD, Chen JT, Koenen K. Racial discrimination, psychological distress, and self-rated health among US-born and foreign-born black Americans. Am J Public Health. 2011;101(9):1704–13.

165. Jasso G, Massey DS, Rosenzweig MR, Smith JP. Immigrant health: selectivity and acculturation. In: Anderson NB, Bulatatoa RA, Cohen B, editors. Critical perspectives on racial and ethnic differences in later life. Washington, DC: National Research Council, National Academies Press; 2003. p. 227–66.

166. Gee GC, Ro A, Shariff-Marco S, Chae D. Racial discrimination and health among Asian Americans: Evidence, assessment, and directions for future research. Epidemiol Rev. 2009;31(1):130–51.

167. Viruell-Fuentes EA, Miranda PY, Abdulrahim S. More than culture: Structural racism, intersectionality theory, and immigrant health. Soc Sci Med. 2012;75(12):2099–106.

168. Karraker MW. Introduction: global migration in the twenty-first century. In: Karraker MW, editor. The other people: interdisciplinary perspectives on migration. New York: Palgrave Macmillan; 2013. p. 3–24.

169. Barbeau EM, Hartman C, Quinn MM, Stoddard AM, Krieger N. Methods for recruiting White, Black, and Hispanic working-class women and men to a study of physical and social hazards at work: the United for Health study. Int J Health Services. 2007;37(1):127–44.

170. Krieger N, Waterman PD, Hartman C, Bates LM, Stoddard AM, Quinn MM, et al. Social hazards on the job: workplace abuse, sexual harassment, and racial discrimination—a study of Black, Latino, and White low-income women and men workers in the United States. Int J Health Services. 2006;36(1):51–85.

171. Quinn MM, Sembajwe G, Stoddard AM, Kriebel D, Krieger N, Sorensen G, et al. Social disparities in the burden of occupational exposures: results of a cross-sectional study. Am J Ind Med. 2007;50(12):861–75.

172. Sembajwe G, Quinn M, Kriebel D, Stoddard A, Krieger N, Barbeau E. The influence of sociodemographic characteristics on agreement between self-reports and expert exposure assessments. Am J Ind Med. 2010;53(10):1019–31.

173. Krieger N, Kaddour A, Koenen K, Kosheleva A, Chen JT, Waterman PD, et al. Occupational, social, and relationship hazards and psychological distress among low-income workers: implications of the "inverse hazard law." J Epidemiol Community Health. 2011;65(3):260–72.

174. Krieger N. Data, "race," and politics: a commentary on the epidemiological significance of California's Proposition 54. J Epidemiol Community Health. 2004;58(8):632–3.

175. Office of Management and Budget. Revisions to the standards for the classification of federal data on race and ethnicity. Federal Registrar Notice. 30 October 1997 [June 17, 2013]. Available from: http://www.whitehouse.gov/omb/fedreg_1997standards/.

176. ADA. Americans with Disabilities Act. 1990 [June 17, 2013]. Available from: http://www.ada.gov/2010_regs.htm.

177. GINA. The Genetic Information Nondiscrimination Act of 2008. 2008 [June 17, 2013]. Available from: http://www.eeoc.gov/laws/statutes/gina.cfm.

178. US Government. Periodic Report of the United States of America to the United Nations Committee on the Elimination of Racial Discrimination concerning the International Convention on the Elimination of All Forms of Racial Discrimination. 2013 [June 27, 2013]. Available from: http://www.state.gov/documents/organization/210817.pdf.

179. US Census. Detailed tables on wealth and ownership assets: 2011. 2013 [June 17, 2013]. Available from: http://www.census.gov/people/wealth/data/dtables.html

180. US Census. People in poverty by selected characteristics: 2010 and 2011. 2013 [June 17, 2013]. Available from: http://www.census.gov/hhes/www/poverty/data/incpovhlth/2011/table3.pdf.

181. Macartney S, Bishaw A, Fontenot K. Poverty rates for selected detail race and Hispanic groups by state and place: 2007–2011: American Community Survey Briefs, 2013. February. ACSBR/11–17:[Available from: http://www.census.gov/prod/2013pubs/acsbr11-17.pdf.

182. US Bureau of Labor Statistics. Labor force characteristics by race and ethnicity, 2011. 2013 [June 17, 2013]. Available from: http://www.bls.gov/cps/cpsrace2011.pdf.

183. National Center for Health Statistics. Health, United States, 2009: with special feature on medical technology. Hyattsville, MD: Centers for Disease Control and Prevention, 2010.

184. Hardy-Fanta C, Lien P-t, Pinderhughes DM, Sierra CM. Gender, race, and descriptive representation in the United States: findings from the Gender and Multicultural Leadership Project. Journal of Women, Politics and Policy. 2006;28(3–4):7–41.

185. Pascoe EA, Richman LS. Perceived discrimination and health: A meta-analytic review. Psychol Bull. 2009;135(4):531–54.

186. Santry HP, Wren SM. The role of unconscious bias in surgical safety and outcomes. Surg Clin North Am. 2012;92(1):137–51.

187. Goto JB, Couto PFM, Bastos JL. Systematic review of epidemiological studies on interpersonal discrimination and mental health. Cadernos de Saúde Pública. 2013;29(3):445–59.

188. Williams DR, Williams-Morris R. Racism and mental health: the African American experience. Ethn Health. 2000;5(3–4):243–68.

189. Williams DR, Neighbors HW, Jackson JS. Racial/ethnic discrimination and health: findings from community studies. Am J Public Health. 2003;93(2):200–8.

190. Schnittker J, McLeod JD. The social psychology of health disparities. Annu Rev Sociol. 2005;31:75–103.

191. Paradies Y. A systematic review of empirical research on self-reported racism and health. Int J Epidemiol. 2006;35(4):888–901.

192. Williams DR, Mohammed SA. Discrimination and racial disparities in health: evidence and needed research. J Behav Med. 2009;32(1):20–47.

193. Brondolo E, Brady ver Halen N, Pencille M, Beatty D, Contrada R. Coping with racism: a selective review of the literature and a theoretical and methodological critique. J Behav Med. 2009;32(1):64–88.

194. Brondolo E, Love EE, Pencille M, Schoenthaler A, Ogedegbe G. Racism and hypertension: a review of the empirical evidence and implications for clinical practice. Am J Hypertens. 2011;24(5):518–29.

195. Brondolo E, Libretti M, Rivera L, Walsemann KM. Racism and social capital: the implications for social and physical well-being. J Soc Issues. 2012;68(2):358–84.

196. Couto PF, Goto JB, Bastos JL. Pressão arterial e discriminação interpessoal: revisão sistemática de estudos epidemiológicos. Arquivos Brasileiros de Cardiologia. 2012;99:956–63.

197. Pachter LM, Coll CG. Racism and child health: a review of the literature and future directions. J Dev Behav Pediatr. 2009;30(3):255–63.

198. Sanders-Phillips K, Settles-Reaves B, Walker D, Brownlow J. Social inequality and racial discrimination: risk factors for health disparities in children of color. Pediatrics. 2009;124(Suppl 3):S176–S86.

199. Wyatt SB, Williams DR, Calvin R, Henderson FC, Walker ER, Winters K. Racism and cardiovascular disease in African Americans. Am J Med Sci. 2003;325(6):315–31.

200. Brondolo E, Rieppi R, Kelly K, Gerin W. Perceived racism and blood pressure: a review of the literature and conceptual and methodological critique. Ann Behav Med. 2003;25(1):55–65.

201. Giscombé CL, Lobel M. Explaining disproportionately high rates of adverse birth outcomes among African Americans: the impact of stress, racism, and related factors in pregnancy. Psychol Bull. 2005;131:662–83.

202. Mays VM, Cochran SD, Barnes NW. Race, race-based discrimination, and health outcomes among African Americans. Annu Rev Psychol. 2007;58(1):201–25.

203. Giurgescu C, McFarlin BL, Lomax J, Craddock C, Albrecht A. Racial discrimination and the black-white gap in adverse birth outcomes: a review. J Midwifery Womens Health. 2011;56(4):362–70.

204. Pieterse AL, Todd NR, Neville HA, Carter RT. Perceived racism and mental health among black American adults: A meta-analytic review. J Couns Psychol. 2012;59(1):1–9.

205. Cuffee Y, Hargraves JL, Allison J. Exploring the association between reported discrimination and hypertension among African Americans: a systematic review. Ethn Dis. 2012;22(4):422–31.

206. Paradies Y. A review of psychosocial stress and chronic disease for 4th world indigenous peoples and African Americans. Ethn Dis. 2006;16(1):295.

207. Nadimpalli SB, Hutchinson MK. An integrative review of relationships between discrimination and Asian American health. J Nurs Scholarsh. 2012;44(2):127–35.

208. Clough J, Lee S, Chae DH. Barriers to health care among Asian immigrants in the United States: a traditional review. J Health Care Poor Underserved. 2013;24(1):384–403.

209. Acevedo-Garcia D, Lochner KA, Osypuk TL, Subramanian SV. Future directions in residential segregation and health research: a multilevel approach. Am J Public Health. 2003;93(2):215–21.

210. Brulle RJ, Pellow DN. Environmental justice: human health and environmental inequalities. Annu Rev Public Health. 2006;27(1):103–24.

211. Mohai P, Pellow D, Roberts JT. Environmental justice. Annu Rev Environ Resour. 2009;34(1):405–30.

212. Kramer MR, Hogue CR. Is segregation bad for your health? Epidemiol Rev. 2009;31(1):178–94.

213. Landrine H, Corral I. Separate and unequal: residential segregation and black health disparities. Ethn Dis. 2009;19(2):179–84.

214. White K, Borrell LN. Racial/ethnic residential segregation: framing the context of health risk and health disparities. Health and Place. 2011;17(2):438–48.

215. White K, Haas JS, Williams DR. Elucidating the role of place in health care disparities: the example of racial/ethnic residential segregation. Health Services Research. 2012;47(3:Part II):1278–1299.

216. Kressin NR, Raymond KL, Manze M. Perceptions of race/ethnicity-based discrimination: a review of measures and evaluation of their usefulness for the health care setting. J Health Care Poor Underserved. 2008;19(3):697.

217. Bastos JL, Celeste RK, Faerstein E, Barros AJD. Racial discrimination and health: A systematic review of scales with a focus on their psychometric properties. Soc Sci Med. 2010;70(7):1091–9.

218. Swanson NG. Working women and stress. J Am Med Womens Assoc. 2000;55(2):76.

219. Raine R. Does gender bias exist in the use of specialist health care? J Health Serv Res Policy. 2000;5(4):237–49.

220. Govender V, Penn-Kekana L. Gender biases and discrimination: a review of health care interpersonal interactions. Background paper prepared for the Women and Gender Equity Knowledge Network of the WHO Commission on Social Determinants of Health. 2007 [June 17, 2013]. Available from: http://www.who.int/social_determinants/resources/gender_biases_and_discrimination_wgkn_2007.pdf.

221. LeResche L. Defining gender disparities in pain management. Clin Orthop Relat Res. 2011;469(7):1871–7.

222. McDonald P. Workplace sexual harassment 30 years on: a review of the literature. Int J Manag Rev. 2012;14(1):1–17.

223. Williamson IR. Internalized homophobia and health issues affecting lesbians and gay men. Health Education Research. 2000;15(1):97–107.

224. Dean L, Meyer I, Robinson K, Sell R, Sember R, Silenzio VB, et al. Lesbian, gay, bisexual, and transgender health: findings and concerns. J Gay Lesbian Med Assoc. 2000;4(3):102–51.

225. Meyer IH. Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. Psychol Bull. 2003;129(5):674–97.

226. Szymanski DM, Kashubeck-West S, Meyer J. Internalized heterosexism: measurement, psychosocial correlates, and research directions. Couns Psychol. 2008;36(4):525–74.

227. De Santis JP. HIV infection risk factors among male-to-female transgender persons: a review of the literature. J Assoc Nurses AIDS Care. 2009;20(5):362–72.

228. Newcomb ME, Mustanski B. Internalized homophobia and internalizing mental health problems: a meta-analytic review. Clin Psychol Rev. 2010;30(8):1019–29.

229. Burns JK. Mental health and inequity: a human rights approach to inequality, discrimination, and mental disability. Health Hum Rights. 2009;11(2):19–31.

230. Thornicroft G, Rose D, Kassam A. Discrimination in health care against people with mental illness. Int Rev Psychiatry. 2007;19(2):113–22.

231. Emerson E, Madden R, Robertson J, Graham H, Hatton C, Llewellyn G. Intellectual and physical disability, social mobility, social inclusion and health: background paper for the Marmot Review Lancaster,

U.K.: Center for Disability Research (CeDR), 2009 [June 17, 2013]. Available from: http://eprints.lancs.ac.uk/26403/1/Disability_Social_Mobility_Social_Inclusion.pdf.

232. Ory M, Kinney Hoffman M, Hawkins M, Sanner B, Mockenhaupt R. Challenging aging stereotypes: strategies for creating a more active society. Am J Prev Med. 2003;25(3, Suppl 2):164–71.

233. Bugental DB, Hehman JA. Ageism: a review of research and policy implications. Soc Issues Policy Rev. 2007;1(1):173–216.

234. Meisner BA. Physicians' attitudes toward aging, the aged, and the provision of geriatric care: a systematic narrative review. Crit Public Health. 2012;22(1):61–72.

CHAPTER 4

INCOME INEQUALITY

Ichiro Kawachi and S. V. Subramanian

WHY CARE ABOUT INCOME INEQUALITY?

Income poverty is bad for health (see Chapter 2). The poor often cannot afford the means to lead a healthy life; for example, assuring adequate nutrition for their children, or paying the utility bills for heating during winter or air-conditioning in the midst of a heat wave. But in addition to being deprived in an absolute sense—that is, the inability to fulfill the basic human needs of food, shelter, and clothing—being poor also means lacking the income to participate fully in society. For example, to be able to participate as a citizen in a wealthy society like America, it is necessary to have access to additional goods and services such as means of communication (the Internet) and transportation (to get to jobs). To paraphrase the Roman Stoic philosopher Seneca the Younger (c. 4 BC–AD 65), to be poor in a wealthy society is the worst kind of poverty.[1] In this chapter, we consider the evidence on whether income inequality is a threat to population health, that is, does the societal distribution of incomes matter for health and well-being over and above being income poor?

Income inequality has risen during the past four decades within many societies, including the United States, prompting some scholars to raise the alarm concerning the corrosive effects of inequality on social cohesion (see for example, Stiglitz [1]). This was not always the case. For instance, within the United States, between the end of World War II up to the Oil Shock of 1973, incomes grew at an equal pace—roughly 2.5% each year—for all household quintile groups across society. Income distribution changed so little during these decades that the academic study of the topic was relegated to a parochial backwater, and at least one economist likened the trends to "watching the grass grow" (2). We fast-forward to 2013, when the Berkeley economist Emmanuel Saez released the latest data on income inequality using data from income-tax records (3). According to Saez—whose data captures the incomes of the richest Americans who are unlikely

1 Lucius Annaeus Seneca. "Occurrent, quod genus egestatis gravissimum est, in divitiis inopes." *Ad Lucilium epistulae morales*: Epistle LXXIV.

to end up on surveys conducted by the Census Bureau—the incomes of the top 0.01% grew by 76.2% during the decade between 2002 and 2012 (i.e., straddling the Great Recession), while the incomes of the bottom 90% *fell* by 10.7% in inflation-adjusted terms. Income inequality in America has now reached the highest levels on record since 1913, when the government first instituted an income tax.

THREE ACCOUNTS LINKING INCOME INEQUALITY TO POPULATION HEALTH

Does income inequality pose a threat to population health? In this chapter we present three distinct accounts of why and how income inequality could be linked to population health outcomes. The three stories are not mutually exclusive; they may all, some, or none of them be correct. We summarize the state of knowledge and empirical evidence supporting each of these accounts.

ABSOLUTE INCOME EFFECT

The three accounts linking income inequality to population health are summarized in Table 4.1 (4, 5). The first account, which we have labeled the **absolute income effect**, hinges on the *shape* of the relationship between individual income and health status. As illustrated in Figure 4.1, the shape of the relationship between individual (or household) income and health is *concave*, that is, the first derivative is positive ($d' > 0$) and the second derivative is negative ($d'' < 0$). Simply stated, it assumes that there are diminishing marginal returns to health from incremental gains in income. The shape of the curve seems to be a robust and almost universally observed feature of the income/health relationship—namely, among households with very low incomes, each

TABLE 4.1: Three explanations linking income inequality to population health

Theory	Formulation	Mechanism
Absolute income effect	$hi = f(yi)$ $f' > 0, f'' < 0$	The concave *shape* of the relationship between income and health predicts that, ceteris paribus, more unequal societies have worse average health.
Relative income effect	$hi = f(yi - yp)$	Income inequality creates a bigger gap between your income and the incomes of others you compare yourself to. The size of this gap leads to stress and frustration.
Contextual effect of income inequality	hi = f (yi, Gini)	When the incomes of the top 1% pull away from the rest, they cause a variety of "pollution effects" on the quality of life the bottom 99%.

Source: Wagstaff and Van Doorslaer (5); and Subramanian and Kawachi (4).

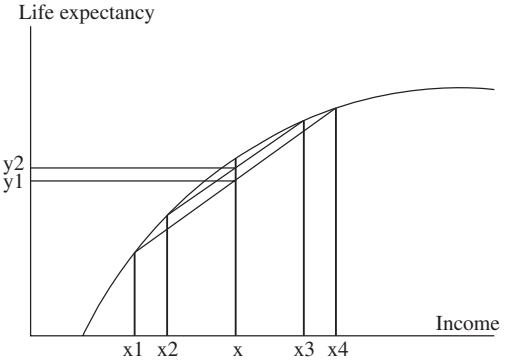


FIGURE 4.1: Theoretical relationship between income and life expectancy.
Source: Rodgers (6).

additional dollar yields a "bigger bang for the buck" than among households that have already sufficient income to fulfill basic needs such as adequate nutrition, shelter, and clothing. Indeed, at some point along the x-axis, the curve must be completely flat because as far as we know there is a theoretical maximum life span (e.g., Genesis 6:3: "And the LORD said, My spirit shall not always strive with man, for that he also is flesh: yet his days shall be an hundred and twenty years"). Thus, the billionaire will not succeed in extending his life span by earning an extra dollar.

As Rodgers (6) noted, the concave shape of the relationship between income and health has an important implication for income distribution and population health. We illustrate this with a thought experiment of two individuals, x1 and x4, living on an otherwise deserted tropical island (Figure 4.1). In the initial condition—*assuming that one's income is causally related to one's life expectancy* (see Chapter 2)—the *average* life expectancy on this imaginary island is predicted to be y1. Now, imagine that we tax the rich person (x4) and transfer the money to the poor person (x1), so that the distribution of income has narrowed from a spread of x1 to x4 to a spread of x2 to x3. The mean income remains the same, that is, we have effected a mean-preserving income transfer. In the post-tax scenario, the predicted average life expectancy on the island has increased to y2. If we generalize the scenario, we can see that among societies with comparable levels of economic development (mean GDP per capita), those with a narrower (i.e., more egalitarian) distribution of income will have higher average life expectancy, all other things being equal. The mechanical reason for this is that any reduction in health as a result of taking away income for the rich is more than offset by the gain in life expectancy for the poor as a result of the transfer.

What Rodgers noticed is simply a restatement of the principle of philanthropy toward health issues. For example, the Giving Pledge is a campaign (started by Warren Buffett and Bill Gates in 2010) to encourage the wealthiest people in the world to donate their wealth to philanthropic causes. When billionaires donate their billions to the world's poor, their generosity is unlikely to adversely impact their life expectancy.[2] On the other hand, when we consider that two-thirds of the planet's population still subsists on incomes below $2 per day, even a few extra dollars channeled

2 Indeed, some evidence suggests that giving away money to help others might improve the health and happiness of the giver as it enables him/her to bask in the glow of doing a good deed (6a). In that case, the health consequences of philanthropy would be positive sum (win/win), not zero sum.

in their direction could mean the difference between life and death, for example, the ability to purchase an insecticide-treated bed net (average cost $3–$5).

In the literature on income inequality and health, some confusion has been sown by referring to the absolute income effect as a "statistical artefact" (7), somehow implying that the association between income distribution and health is spurious. However, there is nothing artifactual about the relationship, so long as we accept that: *(1)* some part of the association between income and health is causal; *(2)* the shape of the relationship between income and health is concave; and *(3)* income transfer from the rich to the poor ***has*** to happen. As Deaton (8) pointed out, the use of the "artifact" terminology is "unfortunate in suggesting that there is no real link between income inequality and health, and that redistributive policy cannot improve average population health. This is far from the case: if income causes health, and if there are diminishing returns, redistribution from rich to poor will improve average population health."

If we accept that income distribution is related to population health via the mechanical shape of the relationship between absolute income and health, how large is the effect? Blakely and Wilson (9) attempted to answer this question for the country of New Zealand. In that country, the researchers linked population census data from 1.3 million working-aged New Zealanders to death records over a three-year follow-up period. This linkage enabled them to empirically estimate the mortality risk at each level of household income (reported by New Zealanders on the census) and then to simulate the consequences of shifting dollars from one part of the income distribution to another. According to the results of this simulation, shifting people's income by 10% toward the mean income (equivalent to a 10% reduction in the Gini coefficient—see Appendix for calculation of the Gini index) would lower total mortality rates in the population by about 4%, adjusting for confounders including age, marital status, education, car access, and neighborhood socioeconomic deprivation score. A 4% reduction in total mortality seems like a modest population impact, but in a country the size of New Zealand it translates to about 1,100 averted deaths each year, which is roughly three times the number of fatal motor vehicle crashes annually ($N \approx 350$). These estimates likely overstate the impact of reducing income inequality because they assume that the full benefit of lowering mortality risk is captured by shifting dollars from the top half of the distribution to the bottom half. Additionally, the estimates do not take into account the potential income losses from implementing the income transfer (Arthur Okun's famous "leaky bucket" problem).[3] Nonetheless, simulations of this kind provide a rough "ballpark" figure of the potential population health penalty paid by society as a result of maintaining high degrees of income inequality.

THE RELATIVE INCOME HYPOTHESIS

The second idea linking income distribution to health is the relative income hypothesis. Here the theory posits that as income inequality rises, it creates a bigger gap between an individual's income and the incomes of others to whom they compare themselves (the term, yi – yp, in Table 4.1). The effect of relative income is thus distinguishable from the effect of absolute income. For instance, an

3 In implementing an income transfer, "the money must be carried from the rich to the poor in a leaky bucket. Some of it will simply disappear in transit, so the poor will not receive all the money that is taken from the rich" (9a: p. 91).

individual may have sufficient income (in an absolute sense) to fulfill basic needs—for example, food and shelter—yet be lacking in the means to purchase goods and services that others in the community are able to afford.

The social comparisons implied by the theory are in turn hypothesized to generate two distinct types of psychological effects: *(1)* positional competition, and *(2)* violations of norms of fairness. Both types of processes are postulated to affect health via stress and frustration. In the case of positional competition—that is, the kind of conspicuous consumption and conspicuous leisure that Thorstein Veblen originally described in the *Theory of the Leisure Class* (10)—the affluent engage in the display of their social status and power via the acquisition of luxury goods and services.[4] However, it is important to recall that positional competition is not just limited to billionaires and their luxury toys, but can be observed even among the middle class and the poor. Hence, the Harvard economist James Duesenberry (11) was among the first to argue that the awareness of the consumption habits of others tends to inspire emulation of these practices—the so-called demonstration effect (or what we call the "keeping up with the Joneses effect"). When those with limited economic means fall prey to demonstration effects—for example, because they become captive to the persuasive efforts of the advertising industry (or what John Kenneth Galbraith [12] termed "want creation")— it can result in dis-saving and debt. In the next section we review empirical evidence from cultural anthropology which documents the health consequences of demonstration effects among the poor. It is also important to add here that demonstration effects are not confined to the consumption of luxury baubles. Many consumption goods that began as luxuries can end up becoming "necessities"—witness, for example, access to high-speed Internet or owning a cellphone. A family in suburban America who cannot afford these items may not be income poor in any absolute sense; yet they would be considered deprived in a relative sense.

From the foregoing description, it is clear that the relative income hypothesis is closely tied to the concept of relative deprivation. As articulated by W. G. Runciman: "We can roughly say that a person is relatively deprived of X (in this instance, X being income) when (i) he does not have X, (ii) he sees some other persons as having X, (iii) he wants X, and (iv) he sees it as feasible that he should have X" (13). Not every individual is sensitive to social comparisons that generate a sense of relative deprivation. For example, consider the choice between two hypothetical scenarios, assuming equal purchasing power parity in both worlds:

A. Living in a world where your current income is $50,000 and everyone else you know earns $25,000; or
B. Living in a world where your current income is $100,000 but everyone else you know earns $250,000

4 For example, Bruce Knecht's book, *Grand Ambition: An Extraordinary Yacht, the People Who Built It, and the Millionaire Who Can't Really Afford It* (2013), can be enjoyed as a contemporary ethnographic study of the lengths to which the affluent engage in positional competition to outdo each other in the construction of luxury yachts. One owner installs a snow-making machine on board, prompting another to build a concert hall to accommodate a 50-member orchestra. Yet another installs a catwalk where supermodels can be flown in to strut the latest fashions, and so on.

Given a choice between these alternatives, surveys have found that about half of respondents opt for (A); that is, they would rather take a lower absolute standard of living so long as they are ahead of everybody else (14). In other words, about half of humanity is sensitive to social comparisons—so much that they would rather trade off higher relative income for lower absolute income. The remaining half—including most trained economists—choose option (B), presumably because they do not care about social comparison and positional competition. Indeed, most mainstream economic models of utility ignore relative considerations, that is, utility is assumed to depend only on the individual's absolute level of endowment (e.g., income) and not on his/her relative position.

Does this mean that half of humanity (those who choose option A) is irrational—or that *Homo economicus* is an inadequate representation of human motivation? As Solnick and Hemenway (14) argue, it may be perfectly "rational" to be concerned about social comparison and positional competition. If everybody else in your community can afford a smartphone for their children but you cannot, this could matter for your child's ability to stay connected to his/her friends, to obtain useful information about homework because the teacher assumes everyone owns a smartphone, and so forth. To dismiss these concerns as "envy" misses the point; there are often real consequences of relative deprivation that go beyond internal feelings of jealousy and shame.

EMPIRICAL EVIDENCE

In the field of population health, two distinct approaches have emerged for empirically testing these ideas: one stems from the field of anthropology and the other from the field of economics. In anthropology, William Dressler (15) and Elizabeth Sweet (16, 17) have pioneered a two-step methodology consisting of: *(1) cultural consensus*, which seeks to establish locally accepted norms of material consumption; and *(2) cultural consonance*, which gauges the degree to which the individual is able to conform to the normative standard of consumption. It is hypothesized that the gap between aspiration and reality will then predict stress-related health outcomes such as blood pressure and depressive symptoms. The anthropological approach thus seeks to define relative deprivation in the space of material *consumption*, as opposed to directly measuring relative income. The relationship between consumption and relative income was originally described by Duesenberry (11), who demonstrated that consumption depends not only on a household's absolute income level, but on its income relative to others. Specifically, Duesenberry argued that a household consumes more when it comes into contact with higher-income households (i.e., the "keeping up with the Joneses" effect). As income inequality expands, one would expect that consumption (even among people in the middle or lower end of the income distribution) to become anchored to an ever-increasing standard set by the top end. For example, houses in America have tended to become bigger over time even as the average family size has shrunk (the "McMansion" phenomenon) (18, 19). By focusing on material consumption, the anthropologist's approach seeks to move beyond *how much money people make* toward capturing *what people do with their money* in order to express their social and symbolic status.

Accordingly, the first stage of establishing "cultural consensus" involves mixed methods (in-depth ethnographic interview of key informants, followed by factor analysis) to establish what the locally shared standard of material consumption is that defines the "successful life." For

example, in rural Brazil, the basket of material goods might include ownership of a television, air conditioner, refrigerator, motorcycle, and so on. In the context of an American suburb, we expect the basket of consumption goods to differ and to include items such as owning the "right" kind of smartphone or designer clothing. In the second stage of the analysis, the researcher establishes "cultural consonance," that is, the extent to which individuals are able to conform to the accepted norms of the "good life." In other words, the goal of the methodology is to quantify the degree of relative deprivation for any given individual within the context of his/her community. As Amartya Sen noted (20), the basket of commodities required to "appear without shame" is less demanding in rural Bangladesh than in suburban America, and hence the methodology of cultural consensus explicitly acknowledges such contextual variation.

In a sample of African-American teenagers in Chicago, Sweet (16) found an interaction between cultural consonance and household socioeconomic status (SES) in predicting the teens' level of blood pressure. Among teens from high-SES backgrounds, the closer they conformed to the cultural consensus, the lower was their blood pressure. By contrast, there was an opposing trend among teens from low SES backgrounds, namely, the more they strove to conform to the cultural norm of material success, the higher was their blood pressure. The result suggests that keeping up with the Joneses is toxic for health, but only when the individual is lacking in the material means to do so with ease. For high-SES individuals, the more conspicuous consumption they engage in, the better they feel—or what Runciman (13) referred to as relative *satisfaction* (which is the opposite of relative deprivation).

In contrast to the anthropologist's approach, the method adopted by economists is to operationalize the concept of relative deprivation by calculating each individual's relative income difference from others in their reference group (21). Under this approach, the social comparisons generated by income inequality are not operationalized in the space of consumption, but rather in terms of the differences in income earned by individuals with similar background characteristics (such as educational attainment or occupation). In other words, the implied mechanism of health effects is via violation of the norm of fairness—that is, receiving the same reward from doing the same work.

The most widely used metric for relative deprivation in the economics field is the Yitzhaki Index, first articulated in a seminal paper by Shlomo Yitzhaki in 1979 (22). According to this approach, the relative deprivation experienced by any given individual *i* with income *yi*, who is part of reference group with N people, can be expressed as:

$$D(y_i) = \int_{y_i}^{y^*} \left[1 - F(z)\right] \mathrm{dz}$$

where y* is the highest income in the reference group, F(z) is the cumulative distribution function, and 1 F(z) is the relative frequency of those with incomes above z (22). For example, if an individual's reference group consists of other coworkers in the same workplace, his/her relative deprivation (RDi) is the sum of differences between his/her income and the incomes of all others in the same workplace who earn more than he/she does, divided by the total number of workers.[5]

5 The division by N is necessary to maintain scale invariance—otherwise the larger one's reference group, the greater would be the degree of relative deprivation experienced by the individual.

It is immediately apparent from this formula that the main challenge inherent in this approach consists of meaningfully defining someone's "reference group." It is unclear that people maintain stable or consistent reference groups in their lives. For example, when Subu leaves his house in the morning, his reference group might consist of other drivers on his commute to work ("Why is everybody else driving a luxury car while I am still driving my ten-year-old Honda?"). At work, his reference group switches to his coworkers ("Why did Ichiro get a bigger bonus this year?"). Finally, back at home his reference group might switch again to the lifestyles depicted on his favorite reality TV show. In short, it is not clear whether individuals carry around a consistent and fixed reference group for making social comparisons.

Nevertheless, economists have attempted to estimate relative deprivation by *assuming* that individuals compare themselves to others with whom they share similar characteristics. In a seminal study of this type, Eibner and Evans (23) calculated the Yitzhaki Index for over 122,000 working-age men enrolled in the National Health Interview Survey. In this study, the researchers used the Yitzhaki formula to calculate individual relative deprivation scores by proceeding *as if* each individual compared their own incomes to other men who were living in the same state, of the same age group, belonging to the same race/ethnic group, and with a similar degree of educational attainment. In other words, the method assumes that a high school drop-out would not be comparing their income to a PhD economist; or that an individual living in Mississippi would not be comparing himself to someone living in Manhattan; or that a 25-year-old entry-level intern would not be comparing himself to a 60-year-old partner in a law firm; and so on. Eibner and Evans then proceeded to estimate the 5-year probability of mortality in a regression model that carefully controlled for absolute income, a vector of covariates (such as age, race, education, marital status), and state-level fixed effects. Across multiple sensitivity analyses, the authors found fairly consistent evidence of an association between relative deprivation and mortality risk. For example, they reported that each 1.0 standard deviation increase in the Yitzhaki Index (based on a reference group defined by age and race) was associated with a 57% excess mortality risk. Using other data, Eibner et al. (24) also reported that relative deprivation was further associated with other stress-related health outcomes, including increased risks of smoking, obesity, and mental health services utilization. Following these initial reports, more recent studies have replicated these findings in diverse societies, including additional studies in the United States (25), Sweden (26), and Japan (27) (for a summary of empirical studies, see Adjaye-Gbewonyo and Kawachi (21).

In summary, a growing number of studies have implicated relative deprivation—and social comparisons based on relative income—as a potential mechanism linking income inequality to health outcomes. Nevertheless, two formidable challenges to empirical demonstrations remain, especially with regard to the Yitzhaki-based approach, namely, *(1)* the difficulty (some would say the impossibility) of establishing a valid reference group for individuals, and *(2)* the collinearity between absolute income and relative income. Regarding the latter, the absolute income level of individuals turns out (not unexpectedly) to be strongly correlated with the degree of relative deprivation, that is, individuals who are lower on the income scale have more people ahead of them; and although the degree of collinearity between relative and absolute income is not perfect, such that only one is identifiable in empirical regressions, there is nonetheless a suspicion of residual confounding. In addition, contrary to Runciman's (13) original theory of relative deprivation, empirical studies based on the Yitzhaki Index have so far failed to demonstrate that as the definition of the reference group is tightened (i.e., more characteristics are piled on as the basis for

social comparison), the risks of adverse health outcomes do not become stronger. This is puzzling and contrary to expectation.

THE CONTEXTUAL EFFECT OF INCOME INEQUALITY

The third and perhaps most controversial explanation linking income inequality to population health is the so-called contextual theory (Table 4.1). This account postulates that over and above the concavity effect of absolute income on health (described earlier), income inequality somehow exerts a "direct" effect on the health of individuals. Some credence for this notion was provided by Wolfson and colleagues (28), who estimated the degree of concavity in the income/mortality relationship using data from the National Longitudinal Mortality Survey. The authors then proceeded to show via simulation that the ecological correlation of state-level income inequality and mortality in the United States is too large to be explained by the concavity effect alone, that is, there must be an additional direct effect of inequality on individual mortality.

The origins of the contextual theory can be traced to a seminal paper by Richard Wilkinson (29), who conjectured that people living in unequal societies end up paying a health "tax." The effect has been likened to a miasma or air pollution, that is, it is difficult for any individual (including even the comfortably well off) to completely escape the deleterious effects of societal inequality (30). What could be the mechanism of such an effect?

In *The Price of Inequality* (1), Joseph Stiglitz advances an argument for how the rent-seeking behavior of the top 1% imposes a tax on the rest of society. The story essentially hinges on the erosion of social cohesion when incomes become massively polarized. The narrative unfolds in two steps. First, as the rich pull away from the rest of society, they literally "secede" from the mainstream of society—by segregating themselves in their own communities (sometimes gated communities that come equipped with 24-hour security), by sending their children away to private schools, by purchasing health services through boutique clinics, by arranging for their private trash collection, and so on. The result is that the rich begin to see less and less reason for why they should be subsidizing everyone else for public services (public education, public hospitals, public libraries) that they do not themselves use. In the second stage of the story, the rich begin to agitate for tax relief. As Stiglitz (1) argues, when power is concentrated in one group, it generally succeeds in getting policies that benefit that group, at the expense of the rest of society. Indeed in the OECD, the countries that have seen the largest increases in the shares of incomes at the top are also the countries that have passed the largest tax cuts to those at the top (31). This narrative neatly encapsulates what has happened in American society during the past two decades with regard to tax policy, regulatory policy, and public investment. Income inequality is therefore viewed as degrading the quality of life for all but the very richest in society. This is a different narrative from the relative deprivation story, because according to the relative income hypothesis, the people who do not care about social comparisons ought not to be adversely affected by rising income inequality. But according to the contextual thesis, even those people might be adversely affected by a deterioration in the quality of public services. As argued by Deaton: "To worry about these consequences of extreme inequality has nothing to do with being envious of the rich and everything to do with the fear that rapidly growing top incomes are a threat to the wellbeing of everyone else" (31).

An additional distinction between the relative income hypothesis and the contextual hypothesis is that the former theory also admits to the possibility that some people at the top might *benefit* from living in an unequal society, that is, a comfortably well-off individual might derive satisfaction from being surrounded by many other people with a lower standard of living (the "big fish in a small pond" effect).[6] Some data support this idea. For example, Kahn et al. (32) found that for Americans in the bottom three quintiles of the income distribution, living in a more unequal area was associated with worse health outcomes. By contrast, for people in the top two quintiles of income, the trend was in the opposite direction—that is, they felt better living in a more unequal area than in a more egalitarian one.

The contextual theory, by contrast, posits that almost everyone in society—except perhaps for the top 1% who can escape to their private island retreats—ends up paying a penalty for living in an unequal society. The pathways linking contextual income inequality to specific health effects remain speculative; indeed, it is an almost universal challenge in social epidemiology to theorize the transition from macro-scale phenomena (such as the societal distribution of income) to micro- or individual-level consequences (in this instance, health outcomes). Nonetheless, we can conceive of different types of "pollution effects" imposed by higher income inequality. For example, Wilkinson and Pickett (33)—as well as Kawachi and Kennedy (19)—have argued that more unequal societies generate more anxiety, shame, depression, and other negative emotions. How does this come about in a society like America?

First of all, many Americans are brought up to believe in meritocracy and social mobility. The prevalence of this belief in American culture is illustrated by the "Horatio Alger myth," referring to the rags-to-riches inspirational stories by the nineteenth- century American author Horatio Alger Jr. (1832 –1899). Cross-national surveys reveal that Americans are much more likely than citizens of other countries to endorse statements such as "People get rewarded for their effort" and, conversely, to disagree with statements such as "Coming from a wealthy family is essential to getting ahead" (34). Contrary to the culturally ingrained myth of the American Dream, however, data reveal that social mobility is lower in America compared to the majority of advanced nations. For example, if we take the intergenerational correlation between the incomes of fathers and sons as one indicator of social mobility, the correlation in the United States (0.47) is much higher (i.e., the society is less socially mobile) than other OECD countries including Norway (0.17), Canada (0.19), Sweden (0.27), Japan (0.34), or France (0.41) (35). There is a strong correlation between income inequality and truncated social mobility. More unequal countries also tend to be ones in which a greater fraction of economic advantage (or disadvantage) is passed on between parents and their children (35). The correlation is probably bidirectional, that is, more inequality hampers social mobility, and truncated mobility generates inequality. Regardless, when we put together these two phenomena—that is, the Horatio Alger myth ("if you strive hard, you will surely succeed") and the reality of truncated mobility in America—we have the ingredients of a potent and toxic combination, in which individuals who strive and fail (as many must), will have no one to blame but themselves. According to Robert Merton's (36) social strain theory, the discrepancy between culturally defined goals (striving for material success) versus the actual

6 For example, there seems to be a growing trend in retirement migration, that is, people from rich countries (such as United States and Japan) moving to poorer countries in their retirement, in order to benefit from higher relative incomes (and cheaper costs of living).

opportunities available to achieve these goals (which are in reality limited) results in anomie, frustration, and maladaptive coping strategies—for example, in the resort to illegitimate means (crime) to attain the goals, or substance abuse as a means of escape from self-blame.

In addition to the psychosocial mechanisms described above, more unequal societies also imply greater exposure to the "pathologies of poverty," including higher rates of crime and violence, and perhaps infectious disease. Another type of spillover effect is the community effects of uninsurance (37). In highly unequal communities in the United States, enormous pressure is exerted on public medical services to treat the indigent and uninsured. This can eventually result in the bankruptcy and closure of local emergency services so that even insured people living in the same communities can end up being denied access. Or even if the ER is somehow able to keep going, the insured may suffer from longer wait times while medical personnel are occupied by the triage of uninsured patients (for whom the local emergency service is the first and only source of care).

One area of debate concerning mechanisms centers on whether the impacts of income inequality on health are mediated primarily via "material" pathways or psychosocial pathways (38). From the "materialist" perspective, it has been argued that a psychosocial interpretation of health inequalities—in terms of perceptions of relative disadvantage and the psychological consequences of inequality—is problematic because it ignores or downplays the structural causes of inequalities. This debate has generated a lot of heat but not much light—not least because it is extremely difficult empirically to tease out psychosocial effects from material effects. For example, the materialists claim that ownership of a car or a house represents possession of important material goods. Yet from a psychosocial perspective, both things also provide a sense of "ontological security" (39), that is, there are important psychological benefits that derive from their ownership, and so far we are not aware of a study design that could tease apart these influences. Nor (we would argue) would it be particularly interesting to do so, since those who hew to a psychosocial interpretation of income inequality have never (to our knowledge) advocated that Prozac should be put into the water supply to make people feel good about inequality. Even if psychosocial processes could explain the adverse health effects of inequality, the solution to the problem would be to fix the structural inequalities of opportunity and investment in society.

EMPIRICAL EVIDENCE

In order to test the contextual theory, what researchers have done is to compare the health outcomes of exchangeable individuals living in communities that differ with regard to the background distribution of incomes. In practice this is accomplished by multilevel regression modeling of the general form:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \alpha_1 \bar{X}_{1j} + \left(u_{0j} + e_{ij}\right)$$

where x_{1ij} references the absolute income level of individual *i* living in the *j*-th community, while estimates the marginal change in the health outcome (y) for a unit change in level of area-level income inequality ($\bar{X}_{1j}$). It is important to control carefully for individual income, since it is a

potential compositional confounding variable, that is, individual income is a common prior cause of both area-level income inequality as well as health outcomes.

In recent years there has been a virtual explosion of multilevel studies investigating the association between area income inequality and health outcomes. Rather than summarize them individually, we refer to a meta-analysis conducted by Kondo and colleagues, who sought to summarize the literature up to 2009 (27). The meta-analysis undertook a systematic search of all relevant databases including PubMed, ISI Web of Science, and the National Bureau for Economic Research. The authors identified 27 multilevel studies—9 of them longitudinal and 18 cross-sectional. There have been dozens more ecological studies addressing the link between income inequality and health, but these were not included in the review.[7] A key result of the random effects meta-analysis was that each 0.05-unit increase in the Gini coefficient was associated with an increase in total mortality rate of 7.8% (95% CI: 5.8 to 9.8%). A 0.05-unit change in the Gini is within the bounds of changes in the Gini actually observed in many countries. It is, for example, roughly equivalent to the increase in Gini index in the United States from 1990 (when it was 0.428) to 2011 when it was 0.477 (40). On the other hand, how big of a deal is a 7.8% excess mortality risk? One view is that compared with the excess risk of mortality associated with income poverty (which can be north of 200%), a 7.8% excess risk seems trivial and a distraction from the more urgent agenda of addressing the needs of the poor. Such a view is based on a misconception of risk. The 7.8% excess risk represents the average across all individuals exposed to higher income inequality (for example, all residents in high-inequality states such as Texas, New York, Louisiana) relative to residents of more egalitarian states such as Wisconsin, Minnesota, and Utah. By contrast, the two-fold excess risk of mortality for income poverty applies to the 15% of households living below the federal poverty threshold. The closest analogy to income inequality derives from studies of the effects of air pollution, where meta-analyses have found that each 10 $\mu g/m^3$ increase in $PM_{2.5}$ particulate air pollution is associated with a 4% increase in all-cause mortality risk (41), that is, in the same ballpark as estimates of the excess mortality risk stemming from the "pollution effects" of income inequality. The air pollution analogy is particularly apt since even the "small" excess risk of 4% was sufficient to prompt the US Environmental Protection Agency to set clean air standards.

FLIES IN THE OINTMENT: CRITIQUES OF THE CONTEXTUAL THEORY

As stated earlier, the contextual theory of income inequality remains the most debated of the three theories linking income distribution to population health. We discuss each of these objections in turn.

7 The problem with ecological studies is that they do not help us to adjudicate between the absolute income hypothesis and the contextual effects hypothesis. Thus, a correlation between area-level Gini and health could be due to either or both of these processes.

WHY HAS LIFE EXPECTANCY CONTINUED TO IMPROVE?

Critics have pointed out that life expectancy has continued to improve for most countries even during the past two to three decades when income inequality has soared. This might seem to be an "ugly fact," potentially fatal to the theory[8]—but we would caution: "Not so fast!" There are at least two challenges in explaining time trends in life expectancy: *(1)* many factors contribute to improvements in health (e.g., advances in medical technology) which may mitigate or mask the effects of adverse influences such as rising inequality; and *(2)* the inverse correlation between rising inequality and improvements in life expectancy over the same time period fails to take account of potential lag effects. For example, concerning the latter point it is well known that cigarette smoking prevalence among US women has been falling during the same period that lung cancer rates have been rising; yet nobody would seriously argue that smoking is protective for lung cancer in women (42). Similarly, female life expectancy has been improved over the same period that obesity rates have risen dramatically, yet few would deny that obesity is a risk factor for excess mortality (even though some may argue about the data for overweight individuals).

To wit, we cannot learn very much from a visual inspection of two trend lines. What is needed is a more rigorous time series analysis linking changes in income inequality with changes in mortality. By doing such an analysis, we may not necessarily find that rising inequality causes a *drop* in life expectancy, due to the offsetting influence of many other contemporaneous factors. The relevant counterfactual is whether the annual gain in life expectancy was slowed down or less than expected based on long-term trends. However, implementing a time series analysis has been hampered by uncertainty over the appropriate lag time between a change in income inequality and changes in health status. Empirical attempts to address this issue have suggested that the strongest "signal" for the health effects of income inequality is found up to a decade out (43). Using the US National Health Interview Survey data 1986–2004 with mortality follow-up data 1986–2006 (n = 701,179), Hui Zheng investigated the lagged effects of national-level income inequality on individual mortality risk (44). These effects were tested by using a discrete-time hazard model where contemporaneous and preceding income inequalities were treated as time-varying person-specific covariates, which then tracked a series of income inequalities that a respondent faced from the survey year until he or she died or was censored. The findings of this analysis suggest that income inequality does not have an instantaneous adverse effect on individual mortality risk, but begins to exert its influence 5 years later, peaking at 7 years, and then diminishing after 12 years (44).

DOES THE HEALTH OUTCOME MATTER?

The question of lag times brings up the issue of the specificity of the association between inequality and health outcomes. Income inequality has been linked to a promiscuous range

8 From Thomas Henry Huxley (1870): "The great tragedy of science—the slaying of a beautiful hypothesis by an ugly fact." (From his Presidential Address at the British Association, "Biogenesis and Abiogenesis;" later published in *Collected Essays*, Vol. 8, p. 229.)

of outcomes, ranging from health behaviors (smoking, obesity, drug use), psychological outcomes (depression, anxiety, self-rated health), and cause-specific mortality (infant mortality, cardiovascular disease, homicide) (33). Researchers frequently turn to whatever health outcome happened to be available in the dataset without specifying the mechanisms. Science would be advanced by clearly stating the mechanism and specifying the hypothesized etiologic period. For example, we might anticipate the lag time between inequality and infant mortality to be quite long—perhaps in the order of decades—given the intergenerational as well as early lifecourse influences on maternal health stock (which in turn determines pregnancy outcomes, birthweight, and infant health). By contrast, we might anticipate that the induction time between inequality and stress-related outcomes (such as mental health problems) might be quite short. Thus, in a longitudinal analysis of the National Epidemiologic Survey on Alcohol and Related Conditions, state-level income inequality at baseline (2001–2002) was associated with an increased risk of *incident* depression (assessed via structured diagnostic interviews) during an average three years of follow-up (45).

An additional question that warrants further investigation is whether there are *cumulative* effects of income inequality on health. For example, when researchers have examined the relationship between poverty and health, they have found significant graded associations between the lifetime number of spells of economic hardship (when incomes fell below 200% of the federal poverty threshold) and measures of physical, cognitive, and psychological functioning (46). Translated to the area of income inequality, the question of whether there are cumulative impacts of "being exposed" to inequality might be fruitfully addressed via multiple-membership models within a multilevel regression framework.

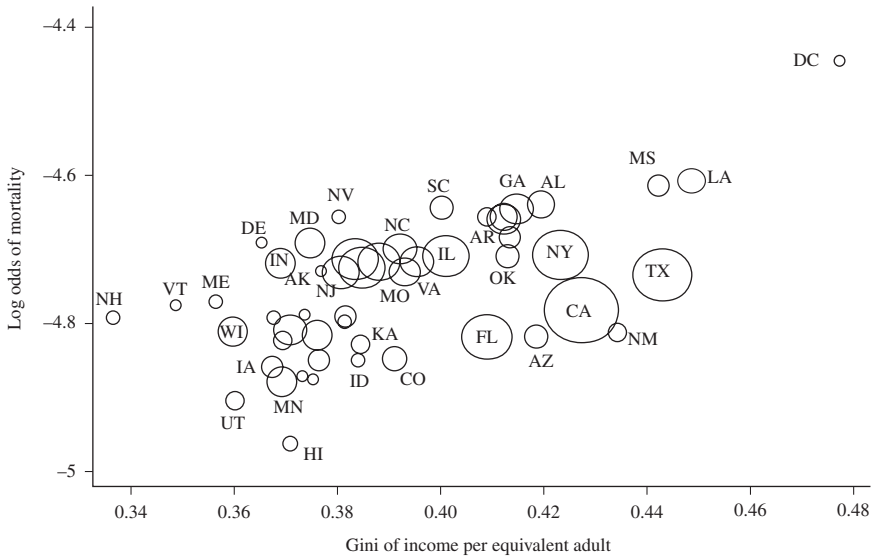
Lastly, the expanded availability of biomarkers in longitudinal datasets provides a potential opportunity for specifying particular pathways through which inequality "gets under the skin" to produce adverse impacts on health. We believe, however, that biomarker measurement is more likely to prove useful in carefully controlled experiments in which income distribution can be manipulated in the laboratory setting.[9] In this kind of set-up, it would be possible to observe the direct consequences of induced inequality on biomarkers of stress, such as cortisol secretion. In a population-based observational study, it is less clear what is the value added by analyzing biomarkers. The reason is because biomarkers (such as markers of inflammation) are affected by health behaviors, such as cigarette smoking. Hence, if income inequality increases frustration and maladaptive coping behaviors (such as more smoking), we *expect* that inflammatory markers would also be elevated in populations exposed to inequality. In that case, the "main story" would be the relationship between inequality and smoking behavior, not the correlation between inequality and the biomarker. In other words, it is sufficient to demonstrate that inequality produces more deleterious behavior without going to the length of mapping the effects on biomarkers. (It is already widely accepted that smoking is deleterious to health, whereas the direct health impacts of elevated inflammatory markers are more open to debate).

9 For an example, see (46a) who experimentally induced inequality in the laboratory by offering different "show up payments" (or honoraria) to the players and observing the effects on cooperative behavior in a trust game.

CONFOUNDING BY RACE

A separate strand of criticism of the contextual theory is that income inequality is not the real culprit driving variations in population health outcomes; rather, it is some other factor which is correlated with it. In the US data, states that have higher levels of income inequality also tend to have greater racial/ethnic heterogeneity (see Figure 4.2, reproduced from Deaton and Lubotsky [47]). Since African Americans have lower incomes compared to white Americans, the higher the proportion of the state population that is black, the higher one would expect income inequality to be. African Americans also suffer from lower life expectancy compared with white Americans (see Chapter 3). Therefore, the association between US state income inequality and mortality may be confounded by the fraction of the population who are black (47). Based on analyses of ecological data, Deaton and Lubotsky (47) reported that *conditional on the fraction black*, neither state-level nor metropolitan area mortality rates are correlated with income inequality. Mortality rates are higher where the fraction black is higher, not only because of the mechanical effect of higher black mortality rates and lower black incomes, but because white mortality rates are higher in places where the fraction black is higher.

The Deaton/Lubotsky critique was subsequently put to the test using multilevel data, that is, data in which race was controlled at both the individual level as well as the state level (as fraction of the population black) for self-rated health as outcome (48) as well as for individual mortality (49). The results of both tests did not suggest that the association between state-level inequality and health is confounded by racial composition. Subramanian and Kawachi (48) analyzed pooled 1995 and 1997 Current Population Surveys, comprising 201,221 adults nested within 50 US states. Controlling for the individual effects of age, sex, race, marital status, education, income, health insurance coverage, and employment status, the authors found a significant effect of state income



Inequality and age-adjusted mortality across US states, 1990 (circles have diameter proportional to population).

FIGURE 4.2: Correlation between US state-level Gini and mortality.

Source: Deaton and Lubotsky (47).

inequality on poor self-rated health. For every 0.05-increase in the Gini coefficient, the odds ratio (OR) of reporting poor health increased by 1.39 (95% CI: 1.26 to 1.51). Additionally controlling for fraction black at the state level failed to explain away the effect of income inequality (OR = 1.30; 95% CI: 1.15 to 1.45). While being black at the individual level was associated with poorer self-rated health, no significant relationship was found between poor self-rated health and the proportion of black residents in a state. In other words, it suggests that in areas of the United States with higher fraction black (e.g., the Southeast), white people experience worse health not because they live in proximity to more black people, but because those areas have higher income inequality.

In an independent test conducted by Backlund and colleagues (49), the US National Longitudinal Mortality Study (NLMS) was used to model the relationship between income inequality in US states and mortality. In multilevel models, the authors adjusted for both race at the individual level and fraction black at the state level. The analyses found that 1990 state-level income inequality was associated with a 22% excess in state-level mortality rates (95% CI: 10 to 37%) for men aged 25–64 and a 5% (95% CI: -6 to 18%) excess risk for women aged 25–64 after adjustment for individual race as well as fraction black. Interestingly, in this analysis, fraction black was also associated with a 14–22% excess risk of mortality (among men and women, respectively) after controlling for state income inequality.

FIXED EFFECTS AND RESIDUAL CONFOUNDING

Although racial composition is not the explanation for the relationship between income inequality and health in the United States, there is nevertheless room for residual confounding by other state-level characteristics. One approach to deal with unobserved confounding is to apply the econometric technique of fixed effects analysis, which cleans out all the time-invariant heterogeneity across the units of observation. In the meta-analysis by Kondo and colleagues (27), there was heterogeneity between the studies that applied fixed effects and those that did not. The pooled relative risk of mortality (for every 0.05-unit increase in the Gini) was 1.016 (95% CI: 0.987–1.046) in the three studies that used the fixed effects approach, that is, the estimate was attenuated close to the null value in contrast to the overall estimate across the studies (1.078). One interpretation of this result is that there is no causal relationship between inequality and health, that is, it reflects unobserved confounding.

This issue has been debated quite extensively in the literature—for example, see the exchanges between Mellor and Milyo (50) and Kawachi and Blakely (42), as well as between Clarkwest (51), Zimmerman (52), and Glymour (53). The crux of the issue is that both cross-country (54) as well as within-country (50, 55) fixed effects analyses rely on examining the impact of a change in income inequality over time on contemporaneous changes in population health (i.e., a first-difference approach). While the motivation for adopting this approach is impeccable (i.e., controlling for all time-invariant unobserved area characteristics), there are some notable limitations of the fixed effects method. First, the time periods involved in testing change-on-change effects may not reflect biologically plausible lag times and induction periods (see earlier discussion about lag times). Second, the approach ignores area differences in intermediary variables (e.g., levels of educational investment) that may reflect the causal effect of Gini in a *prior* time period. That is, by including fixed effects (area dummies), the researcher is controlling for the average differences

across localities in any observable or unobservable predictors of health, such as differences in public investment. However, the very same "time-invariant" variations in area characteristics might also have been caused by income inequality differences between states in a period prior to the start of observation. It is a veritable sledgehammer approach to addressing unobserved heterogeneity in that "methodological concerns regarding unobserved heterogeneity are, of course, entirely valid. [But] the problem lies in taking recourse to models that rely on sources of identification that exclude mechanisms implied by substantive theory" (51).

Last but not least, the fixed effect coefficients soak up all of the between-group "action" so that what is left over is the within-group variation. It is obvious from this that if the exposure shows limited within-group variation over time, we are unlikely to detect the signal. Indeed, in the United States, the between-state variation in income inequality is considerably larger than the within-state changes in income inequality over the kind of time periods examined in fixed effects analyses. In other words, the "dose of treatment" in any given state may not be sufficient to reach a threshold of producing a health effect.

AMERICAN EXCEPTIONALISM?

When we look across all of the empirical studies that have been generated on income inequality, the most robust evidence appears to derive from between-state differences in the United States (30). This is partly a reflection of the availability of better data in the United States. At the same time, we need to move beyond repeated observations of the same natural experiment of N = 50 states (56). But when studies have been carried out in countries beyond the United States, the results have been less consistent. Is this another instance of American exceptionalism, that is, the result of the uniquely toxic combination of the Horatio Alger myth combined with entrenched class immobility in American society?

One clue to this question derives from careful cross-national comparisons of income inequality and health. When Nancy Ross and colleagues (57) contrasted the ecological correlation between income inequality and health in Canada versus the United States, they found no correlation between income inequality and mortality across the 10 Canadian provinces (compared with a strong correlation across 50 US states, such as depicted in Figure 4.2). Tellingly, however, the most unequal Canadian province was still more egalitarian than the most equal US state. This suggests that there may be threshold effects of income inequality on health. Although the correlation between inequality and mortality appears "flat" across the ten Canadian provinces, when we ignore national boundaries and combine Canadian data with US data, all the data points appear to fit along a single regression line.

The presence of a threshold effect of inequality is further buttressed by a cross-national study of income inequality comparing Britain with Japan (58). Across the 30 regions of Britain, Nakaya and Dorling (58) found a steep relationship between the decile ratio of income inequality[10] and working-age standardized mortality ratios (SMR). By comparison, the same relationship across 47 Japanese prefectures appeared flat. Once again, inspection of data from the two countries

10 A summary measure of income inequality which is the ratio of the share of incomes in the bottom 10% of the distribution compared to the top 10%.

reveals that the most unequal Japanese prefecture is more egalitarian than the most equal British region. Just as in the case of the United States and Canada—if we ignore national boundaries—the data from Britain and Japan appear to merge and become part of a greater pattern. The same pattern was reported in an analysis comparing data on metropolitan inequality and mortality from five countries: United States, Britain, Australia, Canada, and Sweden (59). Whereas there was a strong ecological correlation between metropolitan inequality and mortality in the two countries with the highest levels of inequality (United States and Britain), no within-country relationship was found for the three more egalitarian countries—Sweden, Japan, and Australia.

The final piece of evidence in support of a threshold hypothesis stems from repeated tests of the income inequality/health relationship within the same country across different points in time. In Japan (60) and in Taiwan (61), a correlation between inequality and health has emerged in tandem with rising levels of income inequality. In other words, up to a certain level of inequality, there may be little or no adverse health impacts. Indeed, it is possible that flattening the income distribution *too far*—such as happened in formerly Soviet bloc countries—may have led to stifling of individual initiative, proliferation of "under-the-table" forms of compensation (e.g., a dacha in the countryside for those connected to the Communist Party), and consequent feelings of lack of fairness, cynicism, and demoralization. To wit, there may be a "sweet spot" for income inequality—either too much or too little could produce a drag on population health. At present, the number of studies is too few to draw any definitive conclusion about the threshold value of Gini. Nonetheless, in the meta-analysis conducted by Kondo et al. (27), a stronger relationship was detected between income inequality and mortality in countries with Gini ≥ 0.3 (RR = 1.09, 95% CI: 1.07 to 1.12) compared with studies conducted in countries with Gini < 0.3 (RR = 1.02, 95% CI: 0.97 to 1.07).

The meta-analysis by Kondo et al. (27) also provided a more direct test of the American exceptionalism hypothesis. Across the three US-based multilevel studies, the summary relative risk of mortality for each 0.05-unit increase in Gini was 1.06 (95% CI: 1.01 to 1.11). Among the six non-US studies, the corresponding relative risk was 1.09 (95% CI: 1.06 to 1.12). In other words, there is no overall support for the notion that only Americans pay the price of high inequality. Indeed whenever researchers have examined societies that are comparably—or even more—unequal than the United States, they have reported deleterious health impacts of income inequality, for example, in China (62), Chile (63), and Brazil (64).

Lastly, it has been remarked that all tests of the income inequality hypothesis have so far relied on examining the effects of an *increase* in inequality (where data were available for more than one time point). The reason is because across the majority of countries in the world, inequality has steadily risen during the past three decades of increasing globalization (65). The counterfactual trend, that is, observing what happens when a country's income distribution becomes more egalitarian, has proved more difficult to test. Nevertheless, one way of accomplishing this is to examine the health of migrants who move from a more unequal country to a more egalitarian one. Hamilton and Kawachi (66) examined the health of migrants to the United States, leveraging the fact that there are several countries with Gini values that are even higher than the United States (e.g., parts of Latin America and sub-Saharan Africa). Utilizing individual-level data from the March Current Population Survey matched to country-of-origin data on income inequality, the authors found that among immigrants who have resided in the United States between 6 and 20 years, self-reported health—conditional on sociodemographic characteristics—is more

favorable for immigrants who originated from countries that are more unequal than the United States, relative to those who immigrated into the United States from more egalitarian countries.

THE RELEVANCE OF SPATIAL SCALE

So far we have not touched on the issue of the relevant spatial scale at which income inequality is thought to matter for population health outcomes. Empirical tests of the contextual hypothesis have been conducted across almost every conceivable scale of aggregation, from countries, to regions/states/prefectures, to metropolitan areas, down to counties and neighborhoods. One general observation that can be drawn from these studies is that the cross-country relationship between income inequality and health outcomes—especially in industrialized countries—is not robust. Notwithstanding Wilkinson's (29) original demonstration of an ecological correlation between inequality and life expectancy in nine OECD countries that set the whole field in motion, subsequent tests using a broader sample of nations have either failed to replicate Wilkinson's finding (67) or have found that the correlation is attenuated to statistical nonsignificance after controlling for potential confounding factors (50, 68). Based on these conflicting findings, Deaton (8) concluded that there is no evidence that income inequality drives life expectancy and all-cause adult mortality within the industrialized countries.[11] However, he went on to add that the null findings could be driven by inadequate data, particularly in terms of the comparability of cross-country data on income distribution. Conceptual problems of theorizing the linkage between income distribution and health are "dwarfed by measurement problems" (8), and even analyses based on the Luxembourg Income Study (e.g., Judge et al. [67])—which many consider to represent the "gold standard" of comparative analysis of income distribution in selected countries—are not entirely definitive because the country-specific data are "neither fully comparable nor fully accurate" (8). In short, cross-national studies of income inequality and health have been forced to navigate the precarious shoals between good-quality income data in rich countries with restricted variability in income distribution, and a greater range of inequality among poor and middle-income countries with lousy-quality data.

Given the problems of cross-national comparability of income data, researchers have turned their focus on within-country demonstrations of the contextual hypothesis. These have yielded more consistent findings, particularly across states in the United States (30). It is critical when conducting within-country studies that the researcher has a prior theory about the relevance of different geographic jurisdictions. For example, under a federal system such as the United States, there is considerable variation in state generosity of programs such as Medicaid, welfare (TANF), food stamps (SNAP), unemployment benefits, and so forth, all of which may be influenced by the degree of local income inequality, and each of which may contribute to variations in population health outcomes. When we turn to small areas such as neighborhoods, the within-area variance in incomes shrinks relative to the between-area variance because of residential segregation. This means that some very disadvantaged neighborhoods can come out looking quite

11 Deaton does concede that the cross-country relationship between income inequality and infant mortality—at least in poor countries and possibly in rich countries—"is both theoretically plausible and rather better supported by (the admittedly inadequate) data that are available (8: p. 140).

egalitarian (because everyone is equally poor) and at the same time exhibit poor health status (due to poverty).

This is the scenario that plays out in São Paulo, a notoriously segregated city in Brazil, where there is an inverse ecological correlation between district income inequality and mortality, that is, more inequality looks *good* for health (69). In São Paulo, a raw comparison of districts reveals that higher income-inequality areas of the city (Gini ≥ 0.25) have slightly *lower* overall mortality rates compared to lower income-inequality areas (Gini < 0.25). To address this paradox, Chiavegatto Filho et al. (69) developed a propensity model in which they calculated the probability of each neighborhood in São Paulo receiving the "treatment"—in this instance of having a high level of income inequality (Gini ≥ 0.25). Sixteen district-level covariates were used to calculate the propensity score for each neighborhood: presence of *favelas* (slums), poverty rate, median income, education of head of household, household density, proportion of residences with piped water, proportion of residences with garbage collection, proportion of residents without toilet, proportion of heads of household under 21 years old, illiteracy rate of heads of household, illiteracy rate of 8- to 12-year-olds, proportion of teachers per student (5th to 8th grade), HIV/AIDS incidence, proportion of infants (<1 year old), proportion of elderly (>64 years old), and proportion of women. Once the propensity score was calculated for each neighborhood, "exposed" (i.e., high-inequality) neighborhoods were then matched to "unexposed" (low-inequality) neighborhoods and differences in their health outcomes compared. In this approach, each high-inequality neighborhood in São Paulo is matched with another neighborhood based on their potential exchangeability (70). Unmatched neighborhoods were excluded from analysis.

The results showed that prior to propensity score matching, higher income-inequality districts (Gini ≥ 0.25) had slightly lower average mortality compared with low-inequality (Gini < 0.25) neighborhoods; the absolute difference was 2.23 deaths per 10,000 (95% CI: -23.92 to 19.46), which is contrary to the income inequality hypothesis. When the comparison is done after propensity score matching, higher inequality is associated with a significantly higher mortality rate (absolute rate difference: 41.58 deaths per 10,000; 95% CI: 8.85 to 73.3) (69). These data were ecological, however, and they cannot tell us if there are cross-level interactions between income inequality and individual characteristics (such as socioeconomic status). Stated differently, ecological data are silent on the issue of *for whom* inequality is harmful.

FOR WHOM IS INEQUALITY HARMFUL?

Multilevel studies of income inequality have not allowed us to reach any definitive conclusion about whether any specific group (defined by gender, race, socioeconomic status) is particularly vulnerable to the effects of inequality compared with others. The multilevel study based in the National Longitudinal Mortality Study—one of the largest conducted to date—suggested that the excess risk of mortality is restricted to working-age men and women (aged 25 to 64), and that there is no significant association with state-level inequality among Americans over the age of 65 (49). The authors concluded that "this explains why income inequality is not a major driver of mortality trends in the United States because most deaths occur at ages 65 and over" (p. 590).[12]

12 Of course, deaths at younger ages imply more years of potential life lost, and thus it does not take away from the burden of premature mortality associated with income inequality.

However, in the meta-analysis by Kondo et al. (27), there was no statistically significant difference ($p = 0.26$) between the coefficient estimates for populations under the age of 60 years (RR of mortality for each 0.05-unit increase in Gini: 1.06; 95% CI: 1.01 to 1.10) versus populations aged 60 years or over (RR: 1.09; 95% CI: 1.06 to 1.12). In a recent multilevel analysis of older residents (>60 years) residing in São Paulo, Brazil, neighborhood income inequality (measured by the Gini coefficient) was found to be associated with poor self-rated health after controlling for age, sex, income, and education (OR: 1.19; 95% CI: 1.01 to 1.38) (71). An analysis of the data based on structural equation modeling suggested that the adverse effect of income inequality on self-rated health in the elderly is mediated by higher exposure to community violence and lack of physical activity.

A long-standing conjecture is that the adverse impact of income inequality could be more accentuated among the poor compared with the middle class or wealthy (who may be more effectively able to insulate themselves from its corrosive effects). Subramanian and Kawachi (4) systematically examined the interactions between state income inequality, individual poor self-rated health, and a range of individual demographic and socioeconomic markers in the United States, using the 1995 and 1997 Current Population Surveys and the data on state income inequality (represented using the Gini coefficient) from the 1990, 1980, and 1970 US Censuses. This analysis suggested that for a 0.05 change in state income inequality, the odds ratio of reporting poor health was 1.30 (95% CI: 1.17 to 1.45) in a conditional model that included individual age, sex, race, marital status, education, income, and health insurance coverage, as well as state median income. With few exceptions, no strong statistical support for differential effects of state income inequality across different population groups was observed. However, the relationship between state income inequality and poor health was greater for whites compared with blacks (OR =1:34; 95% CI: 1.20 to 1.48) and for individuals with incomes greater than $75,000 compared with less affluent individuals (OR = 1:65; 95% CI: 1.26 to 2.15). The findings, however, primarily suggested an overall (as opposed to differential) contextual effect of state income inequality on individual self-rated poor health.

THE RELATIVE RANK HYPOTHESIS

One last idea that is occasionally invoked in discussions of the health impact of income inequality is the notion that an individual's positional location (or relative rank) in the income hierarchy is a determinant of their health status. Dominance hierarchies have been observed in many animal species (including nonhuman primates), and the individual's rank within the pack has been demonstrated to affect their access to food and their reproductive mates, as well as their physiology and longevity (72). Arguing by analogy, Wilkinson (73) speculated that the adverse physiological consequences of subordinate rank (i.e., occupying a lower socioeconomic position) may be accentuated in a society with greater inequality in the distribution of incomes. Some experimental evidence supports the notion that subordinate rank—as distinct from income differences between individuals—can result in exaggerated stress responses (74). However, caution is warranted in carrying the analogy too far. In the nonhuman primate

literature, for example, whether it is high- or low-ranking animals that are most stressed in a dominance hierarchy turns out to vary as a function of the social organization in different species and populations (72). For example, in some species (such as dwarf mongooses), it is the dominant animals in the troop who suffer the highest indices of physiologic stress, possibly reflecting the demands of constant fighting in order to maintain their dominant position. Yet in other species (such as the savanna baboons and rhesus monkeys), dominance is maintained primarily via psychological intimidation rather than overt physical aggression, and it is the subordinate animals who suffer the most.

Whether the adverse physiological effects of subordinate rank can explain a part of the relation between income inequality and health remains an open question; the proposition is difficult to test empirically, not least because of the high collinearity between absolute income and rank. From an intervention perspective, moreover, if relative rank is what turns out to confer health advantage or disadvantage, then it implies a limited role for social policy in affecting population health. The reason is because policies such as income taxes can change people's absolute and relative incomes, but they will usually preserve rank.

CONCLUSION

In this chapter we have reviewed the three major strands of theory and evidence linking income inequality to population health. The three accounts are not mutually exclusive of each other; they may all be correct (or, depending on the level of skepticism, none of them may be correct). Some scholars do not consider income inequality to be the problem that needs to be fixed; rather, they would argue that the real problem is the underlying political ideology that gave rise to the widening gap between the "haves" and "have nots" (75). According to this view, the maldistribution of income is a by-product, or an epiphenomenon, resulting from broader adversarial class relations, that is, the rich and powerful in society doing bad things to the weak and disenfranchised. Certainly, the rise of neoliberal ideology starting around 1978–1980—with its emphasis on economic liberalization, privatization, deregulation, de-unionization, and the retrenchment of the welfare state—seems to have tracked the global rise in income inequality rather closely (76). At the same time, ideology does not arise in a vacuum. Political and economic philosophy could be viewed as an attempt by the ruling elites to justify the existing social order; that is, the relation between ideology and income inequality is likely to be bidirectional. As income inequality has widened, power has become concentrated at the top, strengthening the ability of the top 1% to preserve their vested interests (1). Growing evidence from social psychology even suggests that the rich are less attuned to the sufferings of the poor; to put it bluntly, they can lack empathy (77). When the incomes of the rich and poor become polarized to the extreme degrees found in American society today, it gives rise to an empathy gap, creating a "careless" society in which some see no contradiction between cutting food stamps for the poor while advocating tax relief for the rich. The extreme concentration of wealth is already viewed (by some) as a threat to economic growth and the functioning of democracy (1). To the growing list of the societal costs of inequality, we have tried to state the case for adding population health as an additional consideration.

REFERENCES

1. Stiglitz J. The price of inequality. New York: Norton; 2012.
2. Aaron HJ. Politics and the professors: the Great Society in perspective. Washington, DC: Brookings Institution Press; 1978.
3. Saez E. Striking it richer: the evolution of top incomes in the United States 2013 [updated 2012 preliminary estimates]. Available from: http://elsa.berkeley.edu/~saez/saez-UStopincomes-2012.pdf.
4. Subramanian S, Kawachi I. Being well and doing well: on the importance of income for health. Int J Social Welfare. 2006;15(Suppl 1):S13–S22.
5. Wagstaff A, Van Doorslaer E. Income inequality and health: what does the literature tell us? Annu Rev Public Health. 2000;21:543–67.
6. Rodgers G. Income and inequality as determinants of mortality: an international cross-section analysis. Popul Stud. 1979;33:343–51.

6a. Dunn EW, Aknin LB, Norton MI. Prosocial spending and happiness: Using money to benefit others pays off. Curr Dir Psychol Sci. 2014;23:41–7.

7. Gravelle H. How much of the relation between population mortality and unequal distribution of income is a statistical artefact? BMJ. 1989;316(7128):382–5.
8. Deaton A. Health, inequality, and economic development. J Econ Perspectives. 2003;41:113–58.
9. Blakely T, Wilson N. Shifting dollars, saving lives: what might happen to mortality rates, and socio-economic inequalities in mortality rates, if income was redistributed? Soc Sci Med. 2006;62(8):2024–34.

9a. Okun A. Equality and efficiency: The big tradeoff. Washington D.C.: The Brookings Institution; 1975.

10. Veblen T. The theory of the leisure class. London, England: Macmillan; 1899.
11. Duesenberry J. Income, saving and the theory of consumption behavior. Cambridge, MA: Harvard University Press; 1949.
12. Galbraith JK. The affluent society. Boston, MA: Houghton Mifflin; 1958.
13. Runciman W. Relative deprivation and social justice. London, England: Routledge & Kegan Paul; 1966.
14. Solnick S, Hemenway D. Is more always better? A survey on positional goods. J Econ Behav Organ. 1998;37(3):373–83.
15. Dressler W, Balieiro M, Ribeiro R, Dos Santos J. A prospective study of cultural consonance and depressive symptoms in urban Brazil. Soc Sci Med. 2007;65(10):2058–69.
16. Sweet E. "If your shoes are raggedy you get talked about": symbolic and material dimensions of adolescent social status and health. Soc Sci Med. 2010;70(12):2029–35.
17. Sweet E. Symbolic capital, consumption, and health inequality. Am J Public Health. 2011;101(2):260–4.
18. Frank R. Luxury fever: why money fails to satisfy in an age of excess. New York: The Free Press; 1999.
19. Kawachi I, Kennedy B. The health of nations: why inequality is harmful to your health. New York: The New Press; 2002.
20. Sen A. Inequality re-examined. Cambridge, MA: Harvard University Press; 1992.
21. Adjaye-Gbewonyo K, Kawachi I. Use of the Yitzhaki Index as a test of relative deprivation for health outcomes: a review of recent literature. Soc Sci Med. 2012;75(1):129–37.
22. Yitzhaki S. Relative deprivation and the Gini coefficient. Q J Econ. 1979;93(2):321–4.
23. Eibner C, Evans W. Relative deprivation, poor health habits, and mortality. J Human Resources. 2005;XL:592–619.
24. Eibner C, Sturn R, Gresenz C. Does relative deprivation predict the need for mental health services? J Ment Health Policy Econ. 2004;7(4):167–75.

25. Subramanyam M, Kawachi I, Berkman L, Subramanian S. Relative deprivation in income and self-rated health in the United States. Soc Sci Med. 2009;69(3):327–34.

26. Aberg Yngwe M, Kondo N, Hägg S, Kawachi I. Relative deprivation and mortality—a longitudinal study in a Swedish population of 4.6 millions, 1990–2006. BMC Public Health. 2012;12:664.

27. Kondo N, Sembajwe G, Kawachi I, Van Dam R, Subramanian S, Yamagata Z. Income inequality, mortality and self-rated health: a meta-analysis of multilevel studies with 60 million subjects. BMJ. 2009;339:b4471.

28. Wolfson M, Kaplan G, Lynch J, Ross N, Backlund E. Relation between income inequality and mortality: empirical demonstration. BMJ. 1999;319(7215):953–5.

29. Wilkinson R. Income distribution and life expectancy. BMJ. 1992;304:165–68.

30. Subramanian S, Kawachi I. Income inequality and health: what have we learned so far? Epidemiol Rev. 2004;26:78–91.

31. Deaton A. The great escape: health, wealth, and the origins of inequality. Princeton, NJ: Princeton University Press; 2013.

32. Kahn R, Wise P, Kennedy B, Kawachi I. State income inequality, household income, and maternal mental and physical health: cross-sectional national survey. BMJ. 2000(321):1311–5.

33. Wilkinson R, Pickett K. The spirit level: why more equal societies almost always do better. London, England: Allen Lane; 2009.

34. Haskins R, Isaacs J, Sawhill I. Getting ahead or losing ground: economic mobility in America. Washington, DC: The Brookings Institute Economic Mobility Project; 2008.

35. Corak M. Income inequality, equality of opportunity, and intergenerational mobility. J Econ Perspectives. 2013;27(3):79–102.

36. Merton R. Social theory and social structure. New York: Free Press; 1957.

37. Institute of Medicine. A shared destiny: the community effects of uninsurance. Washington, DC: National Academies Press; 2003.

38. Lynch J, Smith G, Kaplan G, House J. Income inequality and mortality: importance to health of individual income, psychosocial environment, or material conditions. BMJ. 2000;320(7243):1200–4.

39. Hiscock R, Kearns A, MacIntyre S, Ellaway A. Ontological security and psycho-social benefits from the home: qualitative evidence on issues of tenure. Housing, Theory and Society. 2001;18(1–2):50–66.

40. DeNavas-Walt C, Proctor B, Smith J, Census Bureau. Income poverty, and health insurance coverage in the United States: 2012. Washington, DC: US Government Printing Office; 2013. p. 60–245.

41. Pope 3rd C, Burnett R, Thun M, Calle E, Krewski D, Ito K, et al. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA. 2002;286(9):1132–41.

42. Kawachi I, Blakely T. When economists and epidemiologists disagree. J Health Politics Policy Law. 2001;26:533–41.

43. Blakely T, Kennedy B, Glass R, Kawachi I. What is the lag time between income inequality and health status? J Epidemiol Comm Health. 2000;54(4):318–9.

44. Zheng H. Do people die from income inequality of a decade ago? Soc Sci Med. 2012;75(1):36–45.

45. Pabayo R, Kawachi I, Gilman S. Income inequality among American states and the incidence of major depression. J Epidemiol Community Health. 2014; 68(2):110–5.

46. Lynch J, Kaplan G, Shema S. Cumulative impact of sustained economic hardship on physical, cognitive, psychological, and social functioning. N Engl J Med. 1997;337(26):1889–95.

46a. Anderson LR, Mellor JM, Milyo J. Induced heterogeneity in trust experiments. Exp Econ. 2006;9(3):223–35.

47. Deaton A, Lubotsky D. Mortality, inequality and race in American cities and states. Soc Sci Med. 2003;56(6):1139–53.

48. Subramanian S, Kawachi I. The association between state income inequality and worse health is not confounded by race. Int J Epidemiol. 2003;32(6):1022–8.

49. Backlund E, Rowe G, Lynch J, Wolfson M, Kaplan G, Sorlie P. Income inequality and mortality: a multilevel prospective study of 521 248 individuals in 50 US states. Int J Epidemiol. 2007;36(3):590–6.

50. Mellor J, Milyo J. Reexamining the evidence of an ecological association between income inequality and health. J Health Polit Policy Law. 2001;26(3):487–522.

51. Clarkwest A. Neo-materialist theory and the temporal relationship between income inequality and longevity change. Soc Sci Med. 2008;66(9):1871–81.

52. Zimmerman F. A commentary on "Neo-materialist theory and the temporal relationship between income inequality and longevity change." Soc Sci Med. 2008;66(9):1882–94.

53. Glymour M. Sensitive periods and first difference models: integrating etiologic thinking into econometric techniques: a commentary on Clarkwest's "Neo-materialist theory and the temporal relationship between income inequality and longevity change." Soc Sci Med. 2008;66(9):1895–902.

54. Beckfield J. Does income inequality harm health? New cross-national evidence. J Health Soc Behav. 2004;45(3):231–48.

55. Kravdal O. Does income inequality really influence individual mortality? Results from a "fixed-effects analysis" where constant unobserved municipality characteristics are controlled. Demographic Research. 2008;18:205–32.

56. Subramanian S, Blakely T, Kawachi I. Income inequality as a public health concern: where do we stand? Health Serv Res. 2003;38(1):153–67.

57. Ross N, Wolfson M, Dunn J, Berthelot J, Kaplan G, Lynch J. Relation between income inequality and mortality in Canada and in the United States: cross sectional assessment using census data and vital statistics. BMJ. 2000;320(7239):898–902.

58. Nakaya T, Dorling D. Geographical inequalities of mortality by income in two developed island countries: a cross-national comparison of Britain and Japan. Soc Sci Med. 2005;60(12):2865–75.

59. Ross N, Dorling D, Dunn J, Henriksson G, Glover J, Lynch J, et al. Metropolitan income inequality and working-age mortality: A cross-sectional analysis using comparable data from five countries. J Urban Health. 2005;82(1):101–10.

60. Oshio T, Kobayashi M. Income inequality, area-level poverty, perceived aversion to inequality, and self-rated health in Japan. Soc Sci Med. 2009;69(3):317–26.

61. Chiang T-L. Economic transition and changing relation between income inequality and mortality in Taiwan: regression analysis. BMJ. 1999;319:1162–5.

62. Chen Z, Meltzer D. Beefing up with the Chans: evidence for the effects of relative income and income inequality on health from the China Health and Nutrition Survey. Soc Sci Med. 2008;66(11):2206–17.

63. Subramanian S, Delgado I, Jadue L, Vega J, Kawachi I. Income inequality and health: multilevel analysis of Chilean communities. J Epidemiol Community Health. 2003;57(11):844–8.

64. Pabayo R, Chiavegatto Filho A, Lebrão M, Kawachi I. Income inequality and mortality: results from a longitudinal study of older residents of São Paulo, Brazil. Am J Public Health. 2013;103(9):e43–9.

65. Kawachi I, Wamala S. Poverty and inequality in a globalizing world. In: Kawachi I, Wamala S, editors. Globalization and health. New York: Oxford University Press; 2007.

66. Hamilton T, Kawachi I. Changes in income inequality and the health of immigrants. Soc Sci Med. 2013;80:57–66.

67. Judge K, Mulligan J, Benzeval M. Income inequality and population health. Soc Sci Med. 1998;46(4–5):567–79.

68. Lynch J, Smith G, Hillemeier M, Shaw M, Raghunathan T, Kaplan G. Income inequality, the psychosocial environment, and health: comparisons of wealthy nations. Lancet. 2001;358(9277):194–200.

69. Chiavegatto Filho A, Kawachi I, Gotlieb S. Propensity score matching approach to test the association of income inequality and mortality in Sao Paulo, Brazil. J Epidemiol Community Health. 2012a;66(1):14–7.
70. Oakes JM, Johnson P. Propensity score matching for social epidemiology. In: Oakes JM, Kaufman J, editors. Methods in social epidemiology. San Francisco: Jossey-Bass; 2006.
71. Chiavegatto Filho A, Lebrão M, Kawachi I. Income inequality and elderly self-rated health in São Paulo, Brazil. Ann Epidemiol. 2012b;22(12):863–7.
72. Sapolsky R. The influence of social hierarchy on primate health. Science. 2005;308(5722):648–52.
73. Wilkinson R. Unhealthy societies: the afflictions of inequality. London, England: Routledge; 1996.
74. Mendelson T, Thurston R, Kubzansky L. Affective and cardiovascular effects of experimentally-induced social status. Health Psychol. 2008;27(4):482–9.
75. Coburn D. Income inequality, social cohesion and the health status of populations: the role of neo-liberalism. Soc Sci Med. 2000;51(1):135–46.
76. Harvey D. A brief history of neoliberalism. Oxford: Oxford University Press; 2005.
77. Stellar J, Manzo V, Kraus M, Keltner D. Class and compassion: socioeconomic factors predict responses to suffering. Emotion. 2012;12(3):449–59.

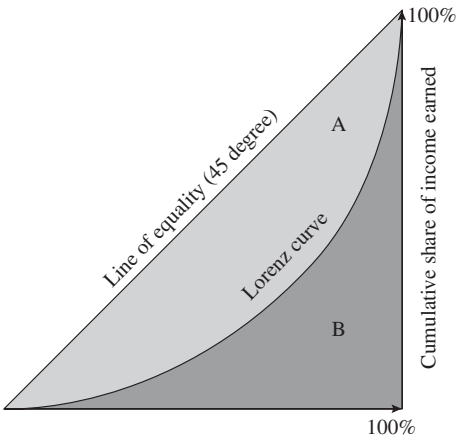
APPENDIX

CALCULATION OF THE GINI COEFFICIENT

The Gini coefficient is a summary measure of income distribution, derived from the Lorenz curve, that plots the proportion of aggregate income (on the vertical axis) earned by each segment of the population, ranked from the poorest to the wealthiest households on the horizontal axis (see diagram). Hypothetically, if every household within a country earned exactly the same income, we would end up plotting a 45-degree line representing perfect equality. For example, the bottom 10% of households would earn 10% of aggregate income; and the bottom half of households would earn half of the total income, and so on. In reality, the distribution of incomes is skewed, so that the bottom 10% of households earns only 5% of total income, and the bottom half of households earns only 30% of total income, and so forth, such that the Lorenz curve dips below the 45-degree line of equality (as depicted in the figure). We thus end up with a visual representation of income inequality—the greater the area between the 45 degree line of perfect equality and the Lorenz curve, the greater the degree of income inequality.

The Gini coefficient is given by the ratio of the area between the line of equality and the Lorenz curve (marked A in the diagram) over the total area under the line of equality (marked A and B in the diagram); namely, Gini = A/(A + B).

From the diagram, it can be seen that the Gini has a theoretical range from 0.0 (perfect equality) to 1.0 (perfect inequality). Under conditions of perfect equality, the Lorenz curve lies on top of the line of perfect equality, and A/(A+B) = 0.0. If one household earned all the income and everybody else earned zero, A/(A+B) becomes 1.0.



Source: Wikipedia (http:||en.wikipedia.org|wiki|File:Economics_Gini_coefficient2.svg).

CHAPTER 5

WORKING CONDITIONS AND HEALTH

Lisa F. Berkman, Ichiro Kawachi, and Töres Theorell

INTRODUCTION

We are living in a rapidly changing world of work, which has profound effects on our health and well-being. These conditions take on particular significance in a globalizing economy. Furthermore, demographic as well as industrial transitions have produced huge changes in the labor force, challenging us to better identify the fit between the workplace conditions and characteristics of the labor force that might impact health. On the demographic front, aging societies with many more older workers, increasing diversity of family composition, and the majority of women joining the paid work force influence the dynamics of labor supply, work opportunities, and work organization. The division of labor between men and women has also changed as cultural expectations, legal challenges to discrimination, and physical job demands have altered how we view "men's" and "women's" work. On the industrial side of the equation, global work forces and outsourcing, the efficiencies of manufacturing jobs, and the transitional nature of jobs themselves all challenge us to rethink how the organization of work may impact the well-being of workers.

We have devoted two chapters in this volume to working conditions. The first, this chapter, covers organizational conditions, practices, and policies that are occurring inside the workplace. Once someone has a job and is experiencing the work environment per se, we are concerned with how it is organized, how flexible the workplace is, and how much job strain, effort/reward, or work-family conflict the worker encounters. We investigate the ways in which opportunities for job or schedule control are structured and how a supervisor's support influences the health of the worker and may even cross over to impact other family members. We then devote a second chapter to work and labor policies at the macro level. Issues of job security, unemployment, and retirement, as well as parental leave and a range of public policies and economic conditions are examined. Roughly, we see these work or labor policies as shaped by forces outside the immediate workplace.

The current chapter presents multiple theoretical frameworks for understanding the ways in which working conditions relate to health. Over the last decade a number of frameworks have been developed and refined, and it is now clear that organizational strains are not unidimensional but that they cross multiple domains. Broadly speaking, workplace influences on health can be conceptualized at three different levels of organization, namely, *(1)* at the level of the job task/job characteristic, *(2)* at the level of the employer/work organization, and *(3)* at the legislative policy level (see Table 5.1).

In this chapter, we focus on the first two of these influences, particularly related to job demands and control and effort/reward, and we discuss new frameworks incorporating aspects of organizational justice, the interactions between the work and home environment (e.g., work/life balance), adverse work schedules and shift work, flexibility, and schedule control. While most frameworks focus on the strains produced by working conditions, it is important to note that over the last several years a number of scientists have called for the study of the *positive and enhancing* aspects of jobs, especially related to social engagement (1–4) and role enhancement, productivity, and meaning (5–10). And, of course, work itself provides not only meaning but also income, so that on the whole work is beneficial to well-being.

Over the last decade, research has focused on the health impacts of six domains of work organization: *(1)* job demands, control, and support; *(2)* effort and reward imbalance; *(3)* organizational justice; *(4)* nonstandard work schedules, including shift and precarious work; *(5)* work and family conflict and associated supervisor and workplace support; and *(6)* schedule control and flexible work arrangements. Many of these models include overlapping elements; for instance, schedule control is a component of many of the models.

Although historically much of the research on workplace strain was started and continues in Europe and the United Kingdom, increasingly, models span continents where economic development, workplace protection, and promotion policies vary widely. To the extent possible, we incorporate an international perspective in this chapter. The chapter begins with a discussion of theoretical models and frameworks and then, within each of the six sections, we turn our attention to the evidence related to the health impacts of each working condition. We conclude with a summary section describing measures and modes of assessment. Here, as in most other chapters, our goal is not to be exhaustive; we draw on several recent meta-analytic reviews and focus on the strongest findings or innovative approaches that could be used more fully in the future. Much of the work in this area focuses on cardiovascular disease, sickness

TABLE 5.1: Hierarchy of workplace influences on health

Level of Organization	Examples of Interventions
Job task/characteristics	Job redesign (to reduce job strain)
Employer/Organization	Workplace health & safety programs Work/life balance
Legislative/Policy	Limits on work hours – e.g., mandatory overtime. Staffing requirements. Limiting piece-rate compensation.

absence, and functional outcomes commonly linked to ergonomic strains, but we also refer to studies with other health outcomes. It is important to note the link between organizational workplace exposures and physically toxic or ergonomically demanding jobs. Often these risks travel together, especially in lower wage jobs. Rarely have investigators been able to tease out the independent effects of physically and socially toxic jobs. There are several excellent recent papers on this topic (see, for instance, 11). We conclude with a discussion of recommendations for future work, and outline public and private policies that have the most potential to improve the work environment, especially for low- and middle-wage workers, women, older workers, and diverse populations least likely to have structured opportunities to balance work and health promotion. The evidence linking many such policies is discussed in greater detail in the next chapter.

HISTORICAL BACKGROUND: INCORPORATION OF JOB STRAIN WITH PHYSICAL HAZARDS IN THE WORKING ENVIRONMENT

Attention to social and organizational aspects of the work environment is a logical outgrowth of the dramatic changes in the workplace. Physical job demands are diminishing for many, though not all, workers, and the growing complexity of modern society increases work-related job demands. While virtually all occupational exposures of importance in public health were related to physical or toxic exposures in the nineteenth and early twentieth centuries (12), starting in the mid-twentieth century it became increasingly clear that the ways in which work was organized could be fundamentally stressful, meaningful, challenging, or demeaning. For instance, analyses of the living conditions of randomly selected working Swedes (13) have shown that noise and heavy lifting are physical conditions of work that have become less frequent during the past 20 years. Containerization and other mechanical processes related to transportation have also changed the physical demands of many jobs in heavily industrialized and developed countries. Job demands, however, in terms of job expectations, multitasking, and interacting with clients, patients, or even other workers, have increased in intensity, according to self-reported data from other studies. In several countries, an increasing number of employees' unions and trade unions, as well as employers, have realized that a functioning social work environment depends heavily on how work is organized and how employees communicate with each other, and that health-promoting factors in the work organization can also improve productivity. In occupational medicine as well, psychosocial working conditions have gained recognition in recent years (14). As women have joined the paid labor force over the last decades and families most often comprise dual wage earners, work-family conflict has been reported to be on the rise among working parents (15, 16). In the United States, some estimates suggest that almost half of working parents experience work-family conflict (17). We note here that workers in several large industries including coal mining, forestry, and agriculture continue to face challenging and risky physical job demands with higher associated workplace

fatalities and disability risks. Several occupational health volumes cover these exposures (18, 19). Our intent here is not to diminish the impact of such risks but rather to extend the range of occupational risks to the social environment.

In the 1960s, as occupational epidemiology grew to include social and psychological experiences of work, several studies examined how working conditions were related to myocardial infarction risk. Many were cross-sectional in nature (20–23). They indicated that there might be a relationship between excessive overtime work and cardiovascular illness risk. Hinkle's study of "night college" men in the Bell Telephone Company was the first prospectively designed study that confirmed an association between excessive demands and myocardial infarction risk (24). Kornitzer and colleagues (25) later observed in a retrospective study of two bank groups in Belgium—one private and one state-owned—that employees in the private banks had a higher incidence of myocardial infarction than employees in the state-owned banks. This difference could not be explained by biomedical risk factors (26). The Belgian bank study was one of the first to indicate a possible relationship between a certain element of working conditions, *work demands* (which were higher in the private banks), and risk of myocardial infarction.

During the 1960s, an important prospective study revealed a higher incidence of myocardial infarction among lower-level as compared to higher-level employees in large companies (27). This evidence raised suspicion for the first time that psychosocial stress might not be a problem primarily for people with a lot of responsibility at the top, as researchers had tended to believe previously, and paved the way for a more complex understanding of the relationship between occupational status and work stress. This new framework revealed that job strain and resulting poor health were most commonly experienced in lower wage and lower status occupations.

WORKING CONDITIONS AND HEALTH

In this section we discuss the theoretical frameworks and links to health outcomes of six dimensions of working life.

THE DEMAND-CONTROL MODEL

One of the most widely used models of job stress is the demand-control model, originally developed by Robert Karasek (28). According to the Karasek model, job strain arises from the interaction of psychological demands with decision latitude (or job control/autonomy). The 2×2 matrix generated by the interaction of psychological demands and job control in turn yields four different types of job conditions.

Active jobs, located in the upper right quadrant, are characterized by high psychological demands *and* a high level of control over managing those demands. Examples of such occupations include physicians, engineers, and other professions. Diagonally opposite to the active quadrant are "passive" jobs, characterized by low demands *and* low control. Examples of passive jobs include janitors and watchmen. These jobs are not psychologically demanding, but, on the other hand, they also lack control; for example, not a lot happens (hopefully) during the typical watchman's shift, but the worker

is not at liberty to abandon his post, and consequently the job is associated with monotony. Over the long term, passive jobs may result in "negative learning," or the gradual atrophy of previously learned skills (29). Thus, the Karasek model acknowledges that not every type of job with a low level of psychological demands (demands that the lay terminology equates with "stress") is healthy; it is the interaction with the amount of control that matters. In the upper left quadrant are "low-strain" jobs, of which a good example is the occupation of social epidemiologist. The most "toxic" type—in the lower right-hand corner—is the "high-strain" job, which combines high demands with low control. Examples of high-strain jobs include assembly line production, as well as many occupations in the modern service economy such as call center operator, waitress, and nurses' aide.

JOB STRAIN AND HEALTH

Sufficient evidence has accumulated during the past two decades to permit meta-analysis of job strain in relation to a variety of health outcomes. Kivimäki and colleagues (30) pooled individual records from 13 European cohort studies (1985–2006) representing 1.49 million person-years of observation. Job strain in these studies was self-assessed through validated job-content questionnaires. The pooled hazard ratio for incident coronary heart disease (adjusted for sex and age) was 1.23 (95% confidence interval [CI]: 1.10, 1.37). In a series of companion papers, the IPD-Work (Individual-participant-data meta-analysis of working populations) Consortium has reported that high job strain was not associated with increased incidence of all-cause cancer (hazard ratio [HR]: 0.97, 95% CI: 0.90, 1.04), or risk of cancer at specific sites, including: colorectal (HR: 1.16, 95% CI: 0.90, 1.48), lung (HR: 1.17, 95% CI: 0.88, 1.54), breast (HR: 0.97, 95% CI: 0.82, 1.14), or prostate (HR: 0.86, 95% CI: 0.68, 1.09) cancers (31).

The meta-analyses of individual data also point to mixed effects of job strain on health-related behaviors. For example, high-strain jobs and passive jobs were associated with sedentarism over follow-up of 2 to 9 years. In a meta-analysis of individual-level data from 14 European cohort studies, the odds of becoming physically inactive during follow-up were about 20% higher for those engaged in high-strain jobs (odds ratio [OR]: 1.21, 95% CI: 1.11, 1.32) as well as for those in passive jobs (OR: 1.20, 95% CI: 1.11, 1.30) (32). However, the temporality of the relationship between job conditions and sedentarism was not unidirectional; that is, physically inactive individuals were also more likely to end up in high-strain jobs over time. With regard to other health behaviors, alcohol consumption showed a U-shaped cross-sectional association with job strain. The analysis defined 4 categories: nondrinkers, moderate, intermediate, and heavy drinkers. Compared to intermediate drinkers, both nondrinkers (random effects OR: 1.10, 95% CI: 1.05, 1.14) and heavy drinkers (OR: 1.12, 95% CI: 1.00, 1.26) had higher odds of job strain, moderate drinkers had lower odds of job strain (OR: 0.92, 95% CI: 0.86, 0.99). There was no clear evidence for longitudinal associations between job strain and alcohol intake (33). Current smokers had higher odds of job strain than never-smokers in cross-sectional data (age, sex, and socioeconomic position-adjusted OR: 1.11, 95% CI: 1.03, 1.18). Current smokers in high-strain jobs smoked, on average, three cigarettes per week more than current smokers without job strain. Once again, however, the analyses of longitudinal data (1 to 9 years of follow-up), showed no clear evidence of an association between job strain and changes in smoking behavior (cessation or initiation) (34). In cross-sectional analyses, relative to normal weight individuals, higher odds of job strain

were found among underweight (OR: 1.12, 95% CI: 1.00, 1.25), obesity class I (OR: 1.07, 95% CI: 1.02, 1.12), and obesity classes II/III participants (OR: 1.14, 95% CI 1.01, 1.28). In longitudinal analysis, both weight gain and weight loss were related to the onset of job strain (35). In summary, the associations between job strain and risky health behaviors (excessive drinking, smoking, overweight/obesity) are stronger cross-sectionally than longitudinally.

There is, however, a modest but robust link between job strain and risk of hypertension. Babu et al. summarized evidence from nine prospective and cross-sectional studies, and found that the pooled odds ratio of hypertension among high-strain workers was 1.3 (95% CI: 1.14, 1.48) for all studies, 3.17 (95% CI: 1.79, 5.60) in case-control studies, and 1.24 (95% CI: 1.09, 1.41) in cohort studies (36). Low decision latitude (or job control) is similarly associated with increased risk of common mental disorders (37).

CURRENT CONTROVERSIES

A major controversy in the demand-control model centers on the question of the "value added" by targeting the reduction of job strain—via job redesign (e.g., boosting workers' autonomy and sense of control)—as a strategy to promote workers' health. In the *Lancet* meta-analysis by Kivimäki and colleagues, 15% of the 197,473 pooled subjects were classified as working in high-job-strain occupations (30). Together with the pooled hazard ratio of 1.23 for incident coronary heart disease, this yields an estimated population attributable risk of around 3.4%.[1] This finding prompted the conclusion that while reducing job strain is an important workplace health promotion goal, it is likely to yield a smaller benefit than attending to other established risk factors, such as cigarette smoking (for which both the prevalence of exposure and the risk of heart disease are likely to be higher).

That conclusion lies at the core of a heated debate about the significance of job strain as a determinant of workers' health. Critics of the *Lancet* report (as well as companion papers from the same IPD-Work Consortium) have argued that the prevalence of job strain was underestimated in the studies that were selected for meta-analysis. For instance, of the 13 cohort studies summarized in the *Lancet* paper, only two represented general working populations, while eight were predominantly among white-collar workers, all of them drawn from European settings (38). The prevalence of high-strain working conditions is greater among blue-collar occupations, especially in rapidly industrializing regions of the world, and hence the true PAR is likely to be higher than 3%. However, even if we assume that the prevalence of high-strain jobs is 50% (i.e., half of the workforce belongs to the bottom right quadrant of Figure 5.1), the estimated PAR is still around 10%.

The other possibility is that the true magnitude of the relative risk of coronary heart disease with job strain is greater than 1.23. One way to check this possibility is via cumulative meta-analysis of existing studies, that is, recomputing the pooled estimate of job strain with the addition of each study in the chronological sequence that they were published. Taking such an approach, Kivimäki et al. (39) showed that the pooled RR was 1.4 (95% CI: 1.0, 1.8) back in 2003 when the first

1 The formula for population attributable risk (PAR) is: P.(RR-1) / P.(RR-1) + 1.

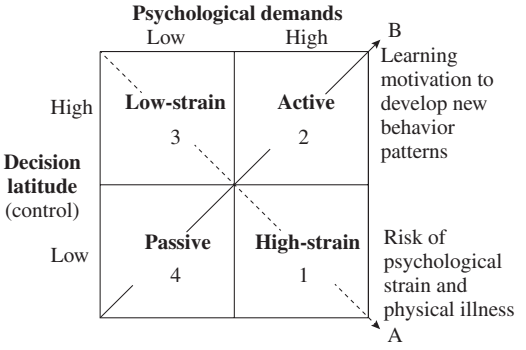


FIGURE 5.1: Psychological Demand-Decision Latitude Model.
Source: Karasek, 1979 (28).

six cohort studies had been published. The subsequent addition of 20 cohort studies did little to change the point estimate. The precision of the estimate improved (due to the addition of more subjects), but the final pooled estimate across 26 cohort studies settled at 1.3 (95% CI: 1.2, 1.5). In other words, the addition of further studies is unlikely to alter the point estimate of the risk by a substantial amount.

What can we take away from this debate? First, as pointed out by the critics of the IPD-Work Consortium meta-analyses, job strain is not the only type of work stressor that matters. As we shall describe below, there are many other stressors in the workplace, including job insecurity (particularly related to nonstandard, precarious, or contingent work), rotating night shifts, organizational injustice, workplace discrimination, bullying, and work-life imbalance. As Landsbergis and colleagues assert, if we add the combined impact of all these work stressors, the PAR is likely to be substantially higher than 3% (40). Secondly, addressing job strain (e.g., via job redesign) does not exclude or compete with health promotion via other means (e.g., worksite smoking cessation). In some instances, the success of health promotion activities in the workplace may be potentiated by attending to the social context of workers' lives. For example, by redesigning jobs to enhance decision latitude, workers' mental health might improve to a point where they are more receptive to messages about quitting smoking, exercising regularly, or eating a healthy diet. Increasingly, health promotion programs are fully integrated with health protection efforts in worker health efforts (41).

Lastly, there is a need for more intervention studies to demonstrate the utility of job redesign as a strategy to improve workers' health. The mixed evidence linking job strain to health behaviors—that is, strong cross-sectional but weak longitudinal associations—raises the distinct possibility that workers with poor health habits *select* themselves into high-strain jobs. One approach to cut the Gordian knot is to conduct experiments manipulating job conditions (preferably through a cluster randomized design) to observe subsequent changes in behavior. One of the few workplace intervention studies that attempted to boost worker autonomy was carried out in a Swedish Volvo plant where the traditional assembly line work (machine-paced, high demand, low control) was replaced by a production process based on a more flexible team-work approach (42). The results of this small intervention found that workers' assessment of autonomy and skill utilization improved, as did their physiological stress profiles, as measured by

epinephrine excretion. Unfortunately, the experiment was not continued for long enough to examine long-term health outcomes; on the other hand, productivity was not adversely affected by the flexible work arrangement, suggesting that job redesign may present a win/win solution for employers and workers.

THE ISO-STRAIN MODEL: EXTENDING THE JOB-STRAIN MODEL TO INCORPORATE SOCIAL SUPPORT AT WORK

An important extension of the demand-control model, proposed by Johnson and Hall, is the iso-strain model, which combines psychological demands and decision latitude with social support from coworkers and supervisors. As described in Chapter 7, on social networks and support, there is an extensive literature on the health benefits of social support. The iso-strain model posits that the most "toxic" jobs are those that combine high strain with social isolation (hence, "iso-strain"). Indeed, many jobs that are characterized by high strain (e.g., call center jobs where workers sit all day in their cubicles) are often associated with low levels of social interactions between coworkers as well as harsh hierarchical relations between supervisors and employees.

Johnson, Stewart, and Hall (43) initially developed the idea that social support at work may affect the relationship between job strain and heart disease. In particular, Johnson's study suggests that people experiencing low social support in conjunction with high psychosocial demands and low control (iso-strain) experience the highest relative risk for cardiovascular disease as compared to the people experiencing other combinations of demand-control and social support (43).

More recently, extended models include workplace support received from multiple sources including supervisors, coworkers, and employing organizations (44; and see Chapter 8, on social capital). Such support can be general or domain-specific: Supervisors can be generally supportive of workers' experiences, in which case they express support in emotional or tangible ways. Workplace support can also be domain-specific; a growing body of research is focused on work-family or work-life conflict. Here (45, 46) supervisor support facilitates effective role functioning in both work and family domains. Hammer, Kossek, and others have extended this model with specific measures of supervisor support that are particular to work-family or work-life demands (46). Recent measures of family-supportive supervisor behaviors (47) show that they impact the physical and mental health of workers, independent of other job-strain measures. Supervisor work-family support is defined as the perception that one's supervisor cares about an employee's work-family well-being, as demonstrated by supervisory helping behaviors to resolve work-family conflict (48), or attitudes such as empathy for work-family balance (45).

THE EFFORT-REWARD IMBALANCE MODEL

In Siegrist's effort-reward imbalance model (6), workers' health is determined by the degree to which workers are rewarded for their efforts. When a high degree of effort does not meet a

high degree of reward, emotional tensions arise and illness risk increases. Effort is defined as the individual's response to the demands made on him or her. These responses may be divided into *extrinsic* effort, which refers to the individual's effort to cope with external demands, and *intrinsic* effort, which corresponds to his or her own drive to fulfill his or her goals. According to Siegrist and colleagues (6, 49), the development of intrinsic effort follows a long-term course in the individual. For example, young employees without extensive work experience with a high degree of "vigor" get involved in more and more commitments. Due to the increasing number of commitments, there may be a rise in the number of conflicts. In turn, overcommitment to work may generate feelings of frustration and irritation. A corporate culture that includes a high level of psychological demands may force employees to internalize extrinsic demands.

Although there is overlap between the effort-reward imbalance and the demand-control-support models, their emphases differ. While the demand-control-support model focuses entirely on the organization's structure, the effort-reward model examines the individual's fit within the environment and includes not only *extrinsic* but also *intrinsic* effort. The latter is closely related to coping—the individual's way of handling difficulties. Reward is a composite measure of financial rewards, self-esteem, and social control. According to the theory, a "healthy state" occurs when reward is increased as effort increases. This state may be achieved by means of external work-related changes such as increased salary and improved social status or increased possibilities for promotion. But it may also be obtained by means of changes in internal effort. Changes in employees' internal effort can come about mainly as a result of changes in coping strategies among the employees, not by altering work conditions per se.

There is some evidence (50) that the decision-latitude component of the demand-control model and the effort-reward imbalance model contribute independently to the prediction of episodes of coronary heart disease. This finding suggests that the models are related to different psychosocial mechanisms linking work conditions to health outcomes. The models share psychological demands (extrinsic effort) but control (decision latitude) and reward are clearly different.

EFFORT-REWARD AND HEALTH

Evidence of the association between effort-reward imbalance and coronary heart disease risk as well as other health outcomes has grown substantially in the last decade. In one of the early studies of blue-collar workers in Germany (49), a very clear relationship was found between effort-reward imbalance and an atherogenic blood lipid pattern. Effort-reward imbalance was also shown to be associated with myocardial infarction risk even after adjusting for accepted biological risk factors (51). Siegrist has summarized findings from several epidemiological studies (6). A recently published study of men and women in the Whitehall Study showed that decision latitude and effort-reward imbalance both contributed independently to the prediction of new events of coronary heart disease in men and women, even after adjustment for a number of biological risk factors and social class (50). Two major reviews by Tsutsumi and Kawakami (52) and van Vegchel, de Jonge, Bosma, and

Schaufeli (53) identify up to 45 studies on effort-reward imbalance. Several recent studies following the publication of these reviews suggest continued interest, and the importance of the concept of effort-reward's independent effects on health and functioning, and its role in explaining socioeconomic gradients in health outcomes. Definitions of the effort-reward model encompass both extrinsic and intrinsic rewards, although current research suggests that extrinsic rewards are central to the operationalization of the effort-reward model. Evidence is strongest for the effort-reward model's association with cardiovascular incidence in men, with findings much weaker or null for women. Workers who report high effort unaccompanied by commensurate rewards were more likely to have cardiovascular morbidity or mortality compared to others. Siegrist et al. report that high-effort-low-reward is positively associated with cardiovascular symptoms or risk factors in the majority (but not all) of studies included in the earlier review (6). A smaller number of studies report associations, inconsistently, with smoking, alcohol consumption, and sickness absence. Again, much of this work has been conducted in exclusively male populations. In a more recent study of a German cohort of male and female workers, Effort-Reward Imbalance (ERI) was associated with HbA1c, an indicator of diabetes risk for men but not for women (54). Recent evidence also found that effort-reward imbalance was modestly associated with functional impairments in an older European cohort of Survey of Health, Ageing and Retirement in Europe (SHARE) participants in a 2-year longitudinal survey (55).

ORGANIZATIONAL JUSTICE

The concept of organizational justice has some overlap with both the demand-control model as well as effort-reward imbalance (56). As originally conceptualized by Moorman (57), organizational justice consists of two domains: distributive justice and procedural justice. Because the definition of distributive justice ("the degree to which a worker believes that she is fairly rewarded in the basis of effort and performance") overlaps to a considerable degree with Siegrist's concept of effort-reward imbalance, researchers have tended to focus on the procedural component of justice. According to Moorman, procedural justice encompasses two dimensions: the existence of formal procedures in the workplace (i.e., the extent to which decision-making processes include input from affected parties, are fair and consistent, and provide useful feedback as well as the possibility of appeal), and interactional justice (the extent to which supervisors treat subordinates with respect, transparency, and fairness). In the public health field, the former dimension has been relabeled "procedural justice," while the latter is referred to as "relational justice." However, as noted by Theorell (58a), the individual items in the procedural justice index overlap with the construct of decision authority in the Karasek demand-control model, while the relational justice scale shares variance with the construct of supervisor support at work.

With these caveats, organizational justice has been shown in a cohort of Finnish workers to be associated with poor self-rated health, minor psychiatric disorders, and sickness absences (58b). In the British Whitehall Study, organizational justice was also shown to predict poor sleep (59) and psychiatric morbidity (60).

WORK-FAMILY CONFLICT: DEMOGRAPHIC TRANSITIONS BRING CHANGE IN WORK LIFE

Beginning in the 1950s and 1960s, women in many countries around the world joined the paid labor force in unprecedented numbers (61). In some countries, fertility declined (either as a cause or consequence of labor force participation) and in other countries (the United States and France, for instance) fertility remained relatively stable, with pregnancies occurring at later ages. While in some countries women struggled to balance work and family demands, in other countries public policies and workplace practices related to parental leave, child care, sickness absence, and tax incentives were enacted to enable working families to have the resources to maintain labor force participation and family life simultaneously.

A second demographic transition has occurred across the globe as a result of declining fertility and decreases in adult mortality influencing population aging. This demographic shift has resulted in fewer young people entering the labor force and has produced an older workforce. Relatively little attention has been paid to the challenges that older workers may face. Finally, the growth of single-parent families over the last decades, along with an increase in both national and cross-national immigration, make family support more difficult to find for many workers with extended families remaining in their native countries. These demographic transitions—while prevalent in richer, industrialized countries—are only now beginning to occur in poorer and less developed countries with a rising set of similar work-related strains. In many developing countries, for instance, one family member must leave his or her rural community to work in urban centers that offer more job opportunities. This immigration—even if temporary—may provide financial security to rural families, but leaves young children and older and sicker family members behind. This demographic phenomenon has contributed to a global rise in work-family conflict.

As developed by Bianchi and others (62, 63) work-family conflict implicitly builds on role theories in which conflicting demands shape strains. At the interface between job and family, both domains can have variable demands and resources or sources of control to moderate those demands. In a measure of work-family conflict developed by Netemeyer (64), conflict can move from work to family, or family to work. Furthermore, in an acknowledgement that spillovers can be both positive and negative, roles can be enhanced as a result of role accumulation. Martikaninen has explicitly identified distinct models of role accumulation (leading to positive benefits) as well as multiple roles (which may be harmful) (65). Bianchi and Milkie (62) further discuss boundary or border theories as being helpful in understanding work-family conflict, as flexibility and permeability shape work and family contexts. For women especially, who even today fulfill the largest obligations with regard to unpaid home care, additional roles in the labor force may lead to exhaustion and illness (66), while also enhancing roles and bringing financial security. Both work and family roles represent core components of adult identity for many men and women, and strains in fulfilling one of these roles brought on by commitments to the other have been hypothesized to cause a host of stress-related outcomes (67–69).

While theories of work-family conflict stand independently of other work-strain theories, there is a way in which they build on the earlier demand, control, support models. Figure 5.2 shows the incorporation of work-family strains into this model to form a single integrated model.

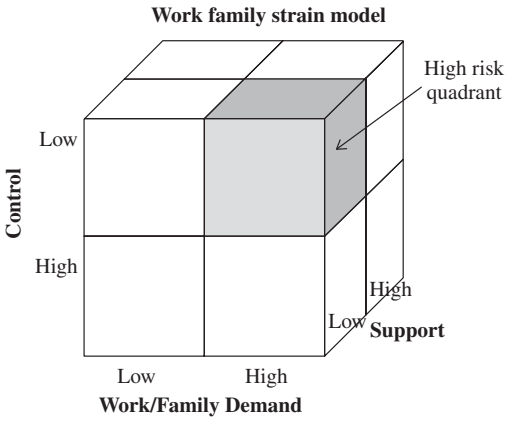


FIGURE 5.2: Theoretical Causal Model of Work-Family Strain.

Source: Berkman LF, O'Donnell EM. The Pro-family Workplace: Social and Economic Policies and Practices and Their Impacts on Child and Family Health—Springer. In NS Landale, SM McHale, A Booth (Eds.), Families and Child Health *(pp. 157–180). University Park, PA: Springer; 2013.*

Our work-family strain model incorporates aspects of family context into the well-established job-strain model discussed earlier. The model links job demands, job control, and social support to a broad range of outcomes for working families, but especially for mothers. Figure 5.2 reveals the three dimensions that create work-family strain. On the left-hand side is the control dimension. On the bottom is work-family demand, and along the third dimension is support. Relevant here is the view that social and economic policies as well as informal family or community support serve as a form of institutional support for families. Families with low control and high demand will be most vulnerable and, in fact, most in need of informal and/or institutional support in the form of public policies. These impacts are further hypothesized to spill over to children. Working women (and to a certain extent men as well) encounter demands from full-time work and high family needs, coupled with low formal support (social protection policies) and often limited informal family support. This combination is exacerbated for low-wage and low-education workers who have little job control and often live near the poverty level. The conflicting demands associated with single parenthood and work-family tensions may affect cardiovascular risks such as smoking and BMI, and likely also affect cardiovascular disease via direct physiological consequences of chronic stress. We hypothesize that the interaction of high work-family demands, low control, and low support leads to sustained stress, damaging health behaviors, and cumulative cardiovascular damage.

Women are especially affected, but men increasingly suffer from many of the same work-family demands. Health effects among mothers may spill over directly to their children during in utero experiences and may play out over childhood via related behavioral and environmental interactions (for example, an employed woman's ability to breastfeed). Variations in work-family strain may arise from differences in family demands, workplace conditions, family protective policies, or informal family supports. Of central interest in Chapter 6 on labor policies is a focus on formal or institutional supports based on family policies.

WORK-FAMILY CONFLICT AND HEALTH: COMPETING DEMANDS OR ENRICHED LIVES

Do men and women differ in the prevalence of work-family conflicts? Or do they have different risks? This is a common theme in seeking to understand the distribution of risk in the population versus the toxicity of that risk. Evidence is mixed with regard to the differences in distribution. In the National Comorbidity Survey, a major study with respect to work-family health findings, Wang et al. (69) report no gender differences in the prevalence of work-family conflict. Nor were there gender differences in risk (e.g., women did not have a higher risk of mental disorders than did men in the study). Rather it is white, well-educated, upper-income middle-aged men and women, and those working over 40 hours a week, who have the highest prevalence of mental disorders. Single women and married men with young children also had higher risks of work-family conflict than others.

Initially, work-family conflict was thought to take its greatest toll on organizational outcomes such as turnover, absenteeism, and job dissatisfaction. In large part, these outcomes were the logical domain of organizational psychologists whose central interest was in the workplace and employee well-being. Since workplace productivity and well-being is closely linked with employee distress and depression, several investigators have linked work-family conflict to psychological distress. As interest in this area grew, a number of social scientists became interested in spillover experiences related to family members—primarily children—who might be affected by parental stress experienced from role overload. Until very recently, however, neither social epidemiologists nor social scientists explored the potential physical health effects such conflict might have on employees themselves. Furthermore, rarely have investigators been able to disentangle the causal direction of associations between work-family conflict and psychological disorders.

Mental disorders are often significantly associated with work-family conflict among both men and women (69), but most studies are essentially cross-sectional, making it unclear whether mental disorders increase the risk of experiencing or reporting work-family conflict. Longitudinal studies, however, support the theory that work-family conflict increases risk for psychological distress and poorer overall well-being (67). In a two-wave longitudinal study conducted at 6 month intervals, work-family conflict predicted self-reported general well-being (70). Work by Frone, Barnes, and Farrell (71) supports the hypothesis that such psychological patterns of distress linked to work-family conflict are further implicated in risky health behaviors including tobacco and alcohol use. Not all studies, however, have confirmed this link; some have suggested more direct links between work-family conditions (both positive and negative aspects of work-family balance) and health behaviors such as alcohol consumption (72, 73). In a study of French Gas and Electricity workers (74), work-family demands were strongly associated with psychiatric sickness absence related both to depression and other psychiatric disorders. The French study is particularly striking because information on psychiatric sickness absence was obtained not from self-report, but from administrative records of actual absences whose causes were certified by a physician.

Information linking work-family conflict to physical health is much scarcer, and is only now beginning to emerge. In a study of long-term care workers, Berkman et al. (75) assessed whether employees in extended care settings with managers who were supportive, open, and

creative about work-family needs (e.g., flexible with work schedules) had lower CVD risk and longer sleep than their less supported counterparts. Employee outcomes were sleep duration (actigraphy) and CVD risk assessed by blood cholesterol, high glycosylated hemoglobin/diabetes, blood pressure/hypertension, body mass index, and tobacco consumption. Employees whose managers were less supportive slept less (29 min/day) and were more than twice as likely to have two or more CVD risk factors (ORs: 2.1 and 2.03 for low and middle manager work-family scores, respectively) than employees whose managers were most open and creative. Employees who provided direct patient care exhibited particularly elevated CVD risk associated with low manager work-family score. In other studies, including the GAZEL Study of French Gas and Electricity workers discussed above, employees with high work and family demands also experienced increased rates of sickness absence for a range of physical disorders (11).

ADVERSE WORK SCHEDULES: SHIFT WORK AND PRECARIOUS WORK

Work schedules characterized by shift work, rotating schedules, and early and late starts are on the rise in much of the world as industries (both service and manufacturing) shift to a 24/7 schedule (76). In addition, involuntary part-time work with few benefits and nonstandard working hours is becoming more common in a globalized economy. In the United States, an estimated 18% of full-time workers spend some time of their workday outside of a 6 am to 6 pm day (76), and a growing number of workers, especially low- and middle-wage workers, have secondary jobs to supplement their earnings. Shift work, characterized by working outside the hours of 8 am to 5 pm, is increasingly common, with night and rotating shift workers experiencing an especially elevated risk for poor health. Rotating shift workers do not work the same shifts over time. While few night shift workers ever make the full shift in circadian rhythms, rotating shift workers can never hope to achieve a full transition, and as a result experience additional health risks (76). Added to the notion of shift work is the experience of early or late start times, with shifts commonly starting before 9 am. Workers in transportation, health care, mining, and construction often have such schedules.

Sleep deficiency is now recognized as one of the most common pathways leading from adverse schedule control to poor health. Sleep deficiency has been shown to be linked to a number of serious health outcomes from metabolic and cardiovascular disorders to accidents and musculoskeletal disorders. Theories linking adverse schedules and working times to health outcomes draw primarily on the biological pathways from sleep deprivation and dysregulation of circadian rhythms to adverse health outcomes. They focus more on the direct physical consequences of such schedules than on the socially mediated mechanisms related to job strain. However, such physiologically "adverse" schedules may simultaneously have both harmful and beneficial social effects. For instance, a number of shift workers select such schedules to maintain family cohesiveness and care where only one parent or caretaker is away from home at a time. At the same time, such shift work may have adverse consequences related to the breakdown of community or family participation. We will review this evidence more fully later in the chapter. Shift work is a special example of psychological and physical demands. As more

industries move to a 24/7 schedule, shift work becomes increasingly common. Shift work entails not only a steady stream of shifts (night shifts, for example) but also often alternating night and day shifts occurring within one weekly or monthly period. Such adverse working hours are physically, socially, and psychologically demanding. At the same time, they increasingly permit dual-earning families to cover child care or other family responsibilities with at least one parent available for most of the day.

NONSTANDARD WORK HOURS AND HEALTH

The evidence linking shift work to adverse health outcomes has grown enormously over the last decade. Vyas et al. (77) reviewed and meta-analyzed 34 studies involving 2,011,935 workers engaged in shift work. Shift work was associated with increased risks of myocardial infarction (risk ratio [RR]: 1.23, 95% CI: 1.15, 1.31) as well as ischemic stroke (1.05, 95% CI: 1.01, 1.09). Several mechanisms have been hypothesized whereby shift work—particularly rotating night shifts—leads to worse health outcomes. Most obviously, shift work disrupts workers' daily routines, so that they end up snacking at odd hours, or becoming socially isolated from their peers. Shift work also suppresses melatonin secretion, leading to increased production of estrogen that may increase the risk for breast cancer. Jia et al. (78) systematically reviewed and meta-analyzed thirteen studies: eight case-control studies and five cohort studies of night work and risk of breast cancer. The pooled RR was 1.20 (95% CI: 1.08, 1.33), although there was heterogeneity between the results of case-control studies (RR: 1.32, 95%CI: 1.17, 1.50) and cohort studies (RR: 1.08, 95%CI: 0.97, 1.21) (78). A major pathway between shift work and adverse health outcomes appears to be sleep disruption (79–81), which in turn influences metabolic function and proinflammatory immune responses, as well as disrupting other physiologic systems. In a study comparing Swedish single-day shift workers and those who worked three shifts, lipid disturbances were identified, though no associations with hyperglycemia were reported (82). More recently, van Mark and colleagues (83) report that findings have not been entirely consistent. In a study of German shift workers no associations were found with IL-6, TNF-alpha, or lymphocyte counts.

Constant rotation between night and day work, usually labeled shift work, is associated with increased risk of developing a myocardial infarction in people of working age (84). Relative risks of the same order as those found for job strain have been found, particularly after many years of exposure to shift work. Knutsson et al. (84) have discussed whether shift work exerts its effect on myocardial infarction risk over and above effects of job strain. On the basis of an extended SHEEP Study, they showed that job strain and shift work were both independently associated with increased myocardial infarction risk after adjustment for accepted biomedical risk factors. More recently, shift work has been linked to a number of chronic diseases (85–87).

PRECARIOUS WORK

The global integration of economies worldwide has resulted in increased pressure for "labor flexibility." A notable aspect of this trend has been the rise in nonstandard work arrangements, which

include (involuntary) part-time work, temporary agency-based work, fixed-term contingent work, and independent contracting (88). Currently, up to one-third of the labor force in industrialized countries is engaged in some form of nonstandard work (sometimes referred to as precarious work). Nonstandard work has a number of advantages, such as the ability of employers to screen workers before hiring them on a permanent basis (thereby lowering training costs), or the ability of workers to control their work schedules (e.g., for juggling caregiving needs at home). The major downside of nonstandard jobs is that they are often correlated with "bad jobs" in the economy: they pay poorly; they lack pension and health benefits; they are insecure; and they are not protected by unions or relevant labor laws. Studies have begun to address the question of whether these "precarious" jobs pose a health hazard for workers. Chapter 6 will review the evidence on job insecurity and workers' health.

FLEXIBILITY AND SCHEDULE CONTROL

The idea that flexibility in working times along with control of schedules may be health-promoting stems from a theory of work-family enrichment or, even more generally, enrichment arising from having multiple roles (89). McNall and colleagues (90) discuss this model of role enrichment relating primarily to the work-family nexus in terms of enriching the rewards, status, and even resource generation that grow from multiple roles. In contrast to theories of work-family conflict or job strain, theories of role enhancement suggest that multiple roles are beneficial for a number of reasons, and that flexibility and schedule control better enable men and women to integrate multiple roles successfully into their lives (90, 91). Greenhaus and Powell (91) define "flexibility" as having discretion in timing, pace, and location of job requirements. This is often translated into flexibility in terms of starting and stopping times at work and offering options for a compressed workweek. "Schedule control" is a closely allied term, preferred by some (92) since "flexibility" can sometimes provide employers with options that make work unpredictable and give less actual control to workers ("just-in-time" staffing, etc.). Such flexibility is closely linked with precarious work conditions and leads to reduced employee control of work hours. Furthermore, schedule control is a domain close to the domains of job control developed by Karasek and Theorell (28, 29) and described above. Schedule control refers specifically to the timing of work, how much people work, when they are able to start and stop work, and whether they can take time off during the workday. Control over work schedules is associated with reduced work-family conflict (93, 94).

A FEW COMMON THREADS: CROSSOVER AND SPILLOVER

Crossover is an interpersonal process by which one person's experiences have implications for others—often this is discussed in work-family frameworks where work strain influences the well-being of children or other family members and is mediated by the "bridging" person's response to strain, which then more directly impacts others. The effects of work strain—whether conceptualized classically as job strain or incorporating new domains of strain—are likely to have crossover effects

on other people in contact with the affected person. So, for instance, children of parents who experience work-family conflict or job strain may be impacted as well and show behavioral and developmental sequelae. Spillover, on the other hand, is most often conceptualized as an intrapersonal experience where an effect on one domain (work life) has an impact on another domain (health or family life). The entire framework for this volume is related to this idea of spillover, that our bodies recognize our social world in highly nuanced and sensitive ways. The concept of embodiment well articulated by Krieger (95) links social, psychological, and biological worlds.

INTEGRATION OF WORKPLACE ORGANIZATION WITH SOCIOECONOMIC CONDITIONS

The nature of social stratification is that the labor market tends to sort workers into different work conditions according to their credentials. Education is the portal to safer, more highly compensated, and prestigious jobs. Conversely, the lack of educational skills and qualifications constrains choice, channeling individuals toward "Dirty, Dangerous and Demeaning" jobs. To the traditional 3Ds, we may add other psychosocial stressors such as low control, job insecurity, and work/life imbalance. In addition, low- and middle-wage earners have more limited family resources, travel longer to their jobs, and have the double burden of physically demanding jobs with ergonomic and/or physically toxic exposures.

One point of debate in the literature concerns whether adverse psychosocial working conditions mediate the relationship between SES and health, or whether work stress is actually a part and parcel of socioeconomic disadvantage. According to Marmot, job strain is a mediator of the relationship between SES and health; that is, one of the pathways through which SES generates health inequalities is via the differential exposure of individuals to psychosocial working conditions (such as job control) (96). In the Whitehall Study, compared with men in the highest grade (administrators), men in the lowest grade (clerical and office support staff) had an age-adjusted odds ratio of 1.50 of developing incident coronary heart disease. For women, the odds ratio in the lowest grade was 1.47 for any CHD. Of the various risk factors for CHD, the authors found that the largest attenuation in the SES gradient was found by adding job control to the regression models; standard coronary risk factors made smaller contributions. Adjustment for all these factors reduced the odds ratios for newly reported CHD in the lowest grade from 1.5 to 0.95 in men, and from 1.47 to 1.07 in women. The implication of this result is that job redesign may be a viable strategy to reduce socioeconomic inequalities in heart disease.

Arguing against this interpretation, Davey Smith and Harding (97) suggested that low control at work is so collinear with low socioeconomic position as to be "virtually synonymous" with each other. They went on to argue that "it is control over the contingencies of life in general, rather than at work in particular, which is important" for the generation of health inequalities (97). There is merit to both sides of the argument. That is, in an occupational cohort, such as the Whitehall Study, it seems plausible that differences in job control could explain a substantial amount of occupational class-based inequalities in health. On the other hand, control over the "contingencies of

life in general" would seem to be an equally important determinant of health, and indeed it is impossible to consider the contribution of work stress in isolation from people's lives outside of their work. Such considerations have underpinned the growing attention to work/life balance, as well as the interaction of the work and home environment.

ASSESSMENT OF THE WORK ENVIRONMENT

In this section, we review measures of *(1)* the demand-control-support model; *(2)* effort-reward balance; *(3)* work-family conflict; and *(4)* work schedules, schedule control, and flexibility. There are a number of modified measures based on a common original measure in many of these cases.

MODES OF ADMINISTRATION

Assessments of the workplace have a long tradition of being multimodal. Specifically, while they are typically self-administered or administered through a direct interview drawing on the respondent's knowledge and perception of workplace conditions, they can also be assessed either through observation or via administrative data. Job matrices for many occupational exposures are based on such observational or administrative data because workers themselves may not be aware of specific toxic exposures. Thus each mode has advantages for specific purposes. Self-administered questionnaires have been used extensively in the study of the psychosocial working environment, mainly because they enable the researcher to conduct studies of large samples efficiently. The disadvantage of such self-report assessments is that they integrate subjective perceptions of workplace experiences with objective working conditions. Interviewer-administered questions have the same issues. Often workplace conditions have been assessed based on observations of specific jobs or industries. Sometimes these observations have then been linked with occupational codes that permit job-strain assessments to be linked with specific occupations. These modes of administration are discussed more fully in appropriate sections.

MEASURING DEMAND-CONTROL

The American Job Content Questionnaire (JCQ) as well as the Swedish version of the demand-control questions have been the most common instruments to assess job demands and job control. The JCQ (available at: http://www.jcqcenter.org/) is a development of the American demand-control-support questionnaire presented in the book *Healthy Work* (29). It is presently used in many countries, whereas the Swedish version is used mainly in Sweden and other Scandinavian countries. The Swedish version has five questions about demands and six about decision latitude. The demand questions deal mainly with quantitative aspects of demands, such as "Do you have time enough to do your work?" and "Do you have to work fast?" but there is also one question that is more qualitative: "Are there conflicting demands in your job?" The decision latitude questions deal both with intellectual discretion (use and development of skills)

and authority over decisions. The questions about intellectual discretion include: "Do you learn new things in your job?", "Is your job monotonous?", and "Does your job require creativity?" The questions about authority over decisions are: "Can you influence how to do your job?" and "Can you influence how your work is to be performed?" The internal consistency of the two dimensions for both men and women in the general working population has proved satisfactory, and factor analysis has confirmed that it is meaningful to group the questions in this way (98). In the American version, there are more questions both about demands and about decision latitude, and there are also several other relevant work dimensions. Its internal consistency has been shown to be satisfactory in several countries. The two versions (which have the same origin—Karasek's initial factor analyses of the American quality of employment surveys in 1968, 1974, and 1977) differ slightly in format since the Swedish version is based on frequency grading (four grades, from "Never" to "Always") of responses to direct questions, and the American one is based on intensity grading (five grades, from "Not at all" to "Very much") of rejection or acceptance of a number of statements.

The operationalization of job strain in these questionnaires has varied. The most frequently used option has been to require demands to be high and decision latitude low at the same time (above/below median or upper/lower quartile or tertile). Another frequently used alternative has been to calculate the ratio between demands and decision latitude and define exposure to job strain equal to those in the upper quartile of this ratio.

MEASURING THE EFFORT-REWARD IMBALANCE MODEL

Siegrist and his collaborators have developed a self-administered questionnaire that includes all the relevant dimensions for the effort-reward imbalance model. A summary of a number of indices built around the effort-reward model has been published with extensive evaluation of a number of domains and short and long forms of the effort reward model (99). Briefly, correlations are very high between partial and complete scales for effort and very high for the reward component. In all cases, complete and partial scales are strongly associated with self-rated health in the expected directions. Sensitivity and specificity of a composite measure, the ratio of effort and reward, suggests that while complete scales have better sensitivity and specificity, partial scales composed of two to three items per domain remain satisfactory.

OTHER MEASURES

Since there are numerous measures of work-family conflict, most commonly built from Netemeyer's original scale, we refer the reader to the references in Table 5.2 for further discussion of these measures. The original Netemeyer scale is designed to reflect the degree to which role responsibilities from one domain are incompatible with the other. Conflict is experienced from work to family as well as family to work. An example of a work-to-family item is: "The demands of your work interfere with your family or personal time." An example of family-to-work conflict is "the demands of your family or personal relationships interfere with work-related activities." Scales have means of item responses ranging from 1 ("strongly disagree") to 5 ("strongly agree").

TABLE 5.2: Measures related to working conditions

1. Demand/Control/Support Model		
Job Demand-Control Model, also called Demand-Discretion Model	(Karasek, 1979 [28]; Karasek & Theorell, 1990 [29])	Assesses psychological strain from the joint effects of the demands related to a work situation and decision-making freedom. Constructed from seven items for job demands and eight measures for job decision latitude (four for decision authority and four for intellectual discretion).
Demand-Control/Support Model	(Johnson & Hall, 1988) (102)	Builds on Karasek's model, adding in the element of social support. Scales constructed from: two items for job demands; 11 items for work control; five items for work-related support.
Demand-Support-Constraint model	(Fletcher & Jones, 1993) (103)	Revision of Karasek's model that addresses several critiques of the model and incorporates additional construct of interpersonal support. Includes four items for job demands and eight items for job decision latitude from Karasek's model as well as four items for interpersonal support.
Job Content Questionnaire and Scale	(Karasek et al., 1998) (104)	Assessment of work quality, building on Karasek's demand-control model. Includes scales for decision latitude, psychological demands, social support, physical demands, and job insecurity. Core version includes 27 questions and the full version has 49 questions.
Pressure Management Indicator	(Williams & Cooper, 1998) (105)	90-item questionnaire assessing job stress. Contains 24 subscales each with multiple items.
2. Effort/Reward Balance		
Effort-reward imbalance model	(Siegrist, 1996 [6]; Siegrist et al., 2004 [106]; Siegrist et al., 2013 [99])	Informed by Karasek's demand-control model and French's person-environment fit model and includes elements of intrinsic effort (coping, need for control), extrinsic effort (demands, obligations), and occupational reward (compensation, esteem, status control). Each of these dimensions is measured through subscales containing several items.

(*continued*)

TABLE 5.2: Continued

3. Work/Family Strain and Conflict		
Model of work, family, and interrole conflict.	(Kopelman, Greenhaus, & Connolly, 1983) (107)	Model includes six variables (work conflict, family conflict, interrole conflict, job satisfaction, family satisfaction, and life satisfaction) measured through a scale with 34 items. Existing measures used for interrole conflict (Pleck et al., 1980) (108) and job satisfaction scales (Hackman and Oldham, 1975) (109).
Coping, social support, and flexibility scales	(Shinn, Wong, Simko, & Ortiz-Torres, 1989) (110)	Includes three dimensions—coping, social support, and job flexibility—that are assessed with separate subscales containing multiple items.
Survey of perceived organizational support	(Eisenberger, Cummings, Armeli, & Lynch, 1997 [111]; Eisenberger, Huntington, Hutchinson, & Sowa, 1983 [112])	36-item scale assessing workers' beliefs about the value placed on their contributions by employers and employers' concern about their well-being.
Work-family conflict scale and family-work conflict scale	(Netemeyer, Boles, & McMurrian, 1996 [64])	Includes separate scales for work-family conflict and family-work conflict, postulating that these are distinct types of interrole conflict. Revision of other scales that do not distinguish between the directionality of conflict (Kopelman et al., 1983 [107], etc.). Each scale includes five items.
Carlson, Kacmar, and Williams scale of work-family conflict	(Carlson, Kacmar, & Williams, 2000) (113)	Building on Netemeyer et al.'s scale, including all three types of work-family conflict (time based, strain based, and behavior based) from both directions of work to family and family to work. The scale contains 20 items.
Negative and positive work-family spillover scales, drawn from the Midlife in the United States (MIDUS) Study	(Grzywacz & Marks, 2000 [72])	Builds on Bronfenbrenner's ecological systems theory (1979) (114). Assesses both work to family and family to work spillovers. Negative spillovers assessed using four items and positive spillovers with three items.
Organizational work-family climate	(Kossek, Colquitt, & Noe, 2001) (115)	Includes separate scales for work climate for family role and family climate for work role, assessed with three items each.

(*continued*)

TABLE 5.2: Continued

Work/family balance measures and supportive supervision	(Clark, 2001) (116)	Includes measures for work culture (temporal flexibility, supportive supervision, and operational flexibility) and work/family balance measures (role conflict, work satisfaction, home satisfaction, family functioning, and employee citizenship).
Family-supportive organization perceptions	(Allen, 2001) (117)	Combines work-family conflict measure from (107), supervisor support from (110), benefit availability and use, and other existing measures for job satisfaction measures, organizational commitment, and intent to turnover.
Work–family enrichment scale	(Carlson, Kacmar, Wayne, & Grzywacz, 2006) (118)	Scale with 18 items assessing three dimensions from the work to family direction (development, affect, and capital) and three dimensions from the family to work direction (development, affect, and efficiency).
Family supportive supervisor behaviors (FSSB) and Family supportive supervisor behavior short-form (FSSB-SF)	(Hammer, Kossek, Bodner, & Crain, 2013 [119]; Hammer, Kossek, Yragui, Bodner, & Hansen, 2009 [48])	Builds on (110), (117), and earlier scales to identify four dimensions of emotional support, role-modeling behaviors, instrumental support, and creative work-family management measured through 28 items.
4. Work Schedules, Schedule Control, and Flexibility		
Job Diagnostic Survey	(Hackman & Oldham, 1975) (109)	Based on a theory that job design affects work satisfaction, the scale contains measures for five core job dimensions (skill variety, task identity, task significance, autonomy, and feedback), critical psychological states, and personal and work outcomes.
Index of work-family policies and the perceived availability of work-family policies and control-time/flex scales	(Eaton, 2003) (120)	Consists of seven items to measure flexibility practices (flextime, part-time jobs, compressed work week, etc.) and seven items to measure perceived availability of flexibility practices as well as measures of formal and informal policies, and a measure of control-time/flex.
Flexible work arrangements	(McNall, Masuda, & Nicklin, 2009 [90])	Assessed through two items. This scale is used in addition to scales for work-family enrichment, job satisfaction, turnover intentions, etc.

FUTURE DIRECTIONS

The workplace is undergoing major changes in almost all parts of the world as economies develop and as different societies adapt to the emerging markets. This dynamic environment makes it an exciting time to study the influences of the workplace on health outcomes. Moreover, it suggests that this may be a time when people can make fundamental changes in the workplace to influence the lives of workers in the future. Simultaneously, demographic transitions relating to *(1)* women joining the paid labor force across the world and *(2)* aging populations leading to both an older workforce and a greater number of people living post retirement for 20 or more years challenge current workplace and retirement practices. Some of the studies we have discussed in this chapter become especially relevant with this end in sight. For example, while there are many observational cohort studies documenting associations between workplace exposures and health outcomes, there are still very few randomized field experiments in which we alter the workplace organization and assess health impacts. This would be a critical next step, and limited evidence suggests that such interventions are not only good for employee health but are linked to increased productivity, lower turnover, and general corporate health as well. Much more of this work needs to be done to assess the practicality of changing the workplace and to strengthen the causal inferences we make about the impacts of working conditions on health. There are several important worksite intervention studies incorporating health as a central outcome. These include the Work, Family and Health Network studies (100) and the NIOSH-supported Total Worker Health program (101), integrating occupational safety and health protection with health promotion to prevent worker injury and illness and to advance health and well-being. In the coming years, results from the large-scale efforts will provide critical information on the health effects of organizational change in worksites.

In these changing times, it will also be important to examine larger social forces that may shape work environments of all people. With this lens, we may better understand how the new work environments will affect health. To date, much of the research in this area has focused predominantly on men, and most of it has been conducted in Europe, where there are numerous social protections in place to improve the well-being of workers in a number of settings. Research that takes such factors into consideration will be well positioned to help make changes in the work environment and to ensure that new work environments are designed to enhance health from the start.

REFERENCES

1. Fried LP, Carlson MC, McGill S, Seeman T, Xue Q-L, Frick K, et al. Experience Corps: a dual trial to promote the health of older adults and children's academic success. Contemp Clin Trials. 2013;36(1):1–13.
2. Berkman L, Ertel K, Glymour MM. Aging and social intervention: life course perspectives. In: Binstock RH, George LK, editors. Handbook of aging and the social sciences. 7th ed. Burlington, MA: Academic Press; 2011. pp. 337–52.
3. Glass TA, Freedman M, Carlson MC, Hill J, Frick KD, Ialongo N, et al. Experience Corps: design of an intergenerational program to boost social capital and promote the health of an aging society. J Urban Health. 2004;81(1):94–105.

4. Reuterwall C, Hallqvist J, Ahlbom A, de Faire U, Diderichsen F, Hogstedt C, et al. Higher relative, but lower absolute risks of myocardial infarction in women than in men: analysis of some major risk factors in the S heep study. J Intern Med. 1999;246(2):161–74.
5. Thoits PA. Multiple identities: examining gender and marital status differences in distress. Am Sociol Rev. 1986;51(2):259–72.
6. Siegrist J. Adverse health effects of high-effort/low-reward conditions. J Occup Health Psychol. 1996;1(1):27–41.
7. Thoits PA. Stress, coping, and social support processes: where are we? What next? J Health Soc Behav. 1995;35(Spec No):53–79.
8. Barnett RC. Women and multiple roles: myths and reality. Harv Rev Psychiatry. 2004;12(3):158–64.
9. Barnett RC, Gareis KC. Parental after-school stress and psychological well-being. J Marriage Fam. 2006;68(1):101–8.
10. Barnett RC, Hyde JS. Women, men, work, and family: an expansionist theory. Am Psychol. 2001;56(10):781–96.
11. Sabbath EL, Glymour MM, Descatha A, Leclerc A, Zins M, Goldberg M, et al. Biomechanical and psychosocial occupational exposures: Joint predictors of post-retirement functional health in the French GAZEL cohort. Adv Life Course Res. 2013;18(4):235–43.
12. Hamilton A. Forty years in the poisonous trades. Am Ind Hyg Assoc J. 1948;9(1):5–17.
13. Statistics Sweden. *The Swedish survey of living conditions.* Design and method. Stockholm: Statistics Sweden; 1996.
14. Gatchel RJ, Schultz IZ, editors. *Handbook of occupational health and wellness.* New York: Springer; 2012.
15. Nomaguchi KM. Change in work-family conflict among employed parents between 1977 and 1997. J Marriage Fam. 2009;71(1):15–32.
16. Duxbury LE, Higgins CA. Work-life balance in the new millennium: where are we? Where do we need to go? Canadian Policy Research Network. Ottawa; 2001. p. 4.
17. Bellavia GM, Frone MR. Work-family conflict. In: Barling J, Kelloway EK, Frone MR, editors. *Handbook of work stress.* Thousand Oaks, CA: Sage Publications; 2005. pp. 113–48.
18. Guidotti TL, Rantanen J, Rose SG, editors. *Global occupational health.* Oxford University Press; 2011.
19. Levy BS, editor. *Occupational and environmental health: recognizing and preventing disease and injury.* Philadelphia: Lippincott Williams & Wilkins, 2006.
20. Biorck G, Blomqvist G, Sievers J. Studies on myocardial infarction in Malmö 1935–1954. II. Infarction rate by occupational group. Acta Medica Scandinavica. 1958;161(1):21–32.
21. Buell P, Breslow L. Mortality from coronary heart disease in California men who work long hours. J Chronic Dis. 1960;11:615–26.
22. Russek HI, Zohman BL. Relative significance of heredity, diet and occupational stress in coronary heart disease among young adults. Am J Med Sci. 1958;235:266–75.
23. Kasanen A, Kallio V, Forrstroem J. The significance of psychic and socio-economic stress and other modes of life in the etiology of myocardial infarction. Ann Med Intern Fenn. 1963;52(Suppl 43):1–40.
24. Hinkle LE, Whitney LH, Lehman EW, Dunn J, Benjamin B, King R, et al. Occupation, education, and coronary heart disease: risk is influenced more by education and background than by occupational experiences, in the Bell System. Science. 1968;161(3838):238–46.
25. Kornitzer M, Kittel F, Dramaix Wilmet M, De Backer G. Job stress and coronary heart disease. Advanced Cardiology. 1982;29:56–61.
26. Kittel F, Kornitzer M, Dramaix M. Coronary heart disease and job stress in two cohorts of bank clerks. Psychother Psychosom. 1980;34(2–3):110–23.

27. Pell S, d'Alonzo CA. Acute myocardial infarction in a large employed population: report of six-year study of 1,356 cases. JAMA. 1963;185:831–41.

28. Karasek RA. Job demands, job decision latitude, and mental strain: Implications for job redesign. Administrative Science Quarterly. 1979;24(2): 285–308.

29. Karasek R, Theorell T. *Healthy work.* New York: Basic Books; 1990.

30. Kivimäki M, Nyberg ST, Batty GD, Fransson EI, Heikkilä K, Alfredsson L, et al. Job strain as a risk factor for coronary heart disease: a collaborative meta-analysis of individual participant data. Lancet. 2012;380(9852):1491–7.

31. Heikkilä K, Nyberg ST, Theorell T, Fransson EI, Alfredsson L, Bjorner JB, et al. Work stress and risk of cancer: meta-analysis of 5700 incident cancer events in 116,000 European men and women. BMJ. 2013;346:f165–5.

32. Fransson EI, Heikkilä K, Nyberg ST, Zins M, Westerlund H, Westerholm P, et al. Job strain as a risk factor for leisure-time physical inactivity: an individual- participant meta-analysis of up to 170,000 men and women. Am J Epidemiol. 2012;176(12):1078–89.

33. Heikkilä K, Nyberg ST, Fransson EI, Alfredsson L, De Bacquer D, Bjorner JB, et al. Job strain and alcohol intake: a collaborative meta-analysis of individual-participant data from 140,000 men and women. PLoS ONE. 2012;7(7):e40101.

34. Heikkilä K, Nyberg ST, Fransson EI, Alfredsson L, De Bacquer D, Bjorner JB, et al. Job strain and tobacco smoking: an individual-participant data meta-analysis of 166,130 adults in 15 European studies. PLoS ONE. 2012;7(7):e35463.

35. Nyberg ST, Heikkilä K, Fransson EI. Job strain in relation to body mass index: pooled analysis of 160,000 adults from 13 cohort studies. J Intern Med. 2012;272:65–73.

36. Babu GR, Jotheeswaran AT, Mahapatra T, Mahapatra S, Kumar A Sr, Detels R, et al. Is hypertension associated with job strain? A meta-analysis of observational studies. Occup Environ Med. 2013;71(3): 220–7.

37. Stansfeld S, Candy B. Psychosocial work environment and mental health: a meta-analytic review. Scand J Work Env Hea. 2006;32(6):443–62.

38. Choi B, Schnall P, Ko S, Dobson M, Baker D. Job strain and coronary heart disease. Lancet. 2013;381(9865):448.

39. Kivimäki M, Ferrie JE, Kawachi I. Cumulative meta-analysis of job strain and coronary heart disease: implications for future research. Am J Epidemiol. 2013;177(1):1–2.

40. Landsbergis PA, Dobson M, Schnall P. RE: Need for more individual-level meta-analyses in social epidemiology: example of job strain and coronary heart disease. Am J Epidemiol. 2013;178(6):1008–9.

41. Sorensen G, McLellan D, Dennerlein JT, Pronk NP, Allen JD, Boden LI, et al. Integration of health protection and health promotion. J Occup Environ Med. 2013;55(12 Suppl):S12–8.

42. Melin B, Lundberg U, Söderlund J, Granqvist M. Psychological and physiological stress reactions of male and female assembly workers: a comparison between two different forms of work organization. J Organiz Behav. 1999;20(1):47–61.

43. Johnson JV, Stewart W, Hall EM. Long-term psychosocial work environment and cardiovascular mortality among Swedish men. Am J Public Health. 1996;86(3):324–31.

44. Kossek EE, Pichler S, Bodner T, Hammer LB. Workplace social support and work-family conflict: a meta-analysis clarifying the influence of general and work-family specific supervisor and organizational support. Pers Psychol. 2011;64(2):289–313.

45. Thomas LT, Ganster DC. Impact of family-supportive work variables on work-family conflict and strain: a control perspective. J Appl Psychol. 1995;80(1):6.

46. Hammer LB, Neal MB, Newsom JT, Brockwood KJ, Colton CL. A longitudinal study of the effects of dual-earner couples' utilization of family-friendly workplace supports on work and family outcomes. J App Psychol. 2005;90(4):799–810.

47. Hammer LB, Kossek EE, Anger WK, Bodner T, Zimmerman K. Clarifying work-family intervention process: the roles of work-family conflict and family supportive supervisor behaviors. J Appl Psychol. 2011;96(1):134–50.

48. Hammer L, Kossek E, Yragui N, Bodner T, Hansen G. Development and validation of a multi-dimensional scale of family supportive supervisor behaviors (FSSB). Journal Manage. 2009;35:837–56.

49. Siegrist J, Matschinger H, Cremer P, Seidel D. Atherogenic risk in men suffering from occupational stress. Atherosclerosis. 1988;69(2–3):211–8.

50. Bosma H, Peter R, Siegrist J, Marmot M. Two alternative job stress models and the risk of coronary heart disease. Am J Public Health. 1998;88(1):68–74.

51. Siegrist J, Peter R, Junge A, Cremer P, Seidel D. Low status control, high effort at work and ischemic heart disease: prospective evidence from blue-collar men. Soc Sci Med. 1990;31(10):1127–34.

52. Tsutsumi A, Kawakami N. A review of empirical studies on the model of effort–reward imbalance at work: reducing occupational stress by implementing a new theory. Soc Sci Med. 2004;59(11):2335–59.

53. van Vegchel N, de Jonge J, Bosma H, Schaufeli W. Reviewing the effort–reward imbalance model: drawing up the balance of 45 empirical studies. Soc Sci Med. 2005;60(5):1117–31.

54. Li J, Jarczok MN, Loerbroks A, Schöllgen I, Siegrist J, Bosch JA, et al. Work stress is associated with diabetes and prediabetes: cross-sectional results from the MIPH industrial cohort studies. Int J Behav Med. 2012;20(4):495–503.

55. Reinhardt JD, Wahrendorf M, Siegrist J. Socioeconomic position, psychosocial work environment and disability in an ageing workforce: a longitudinal analysis of SHARE data from 11 European countries. Occup Environ Med. 2013;70(3):156–63.

56. Kawachi I. Injustice at work and health: causation or correlation? Occup Environ Med. 2006;63(3):578–9.

57. Moorman RH. Relationship between organizational justice and organizational citizenship behaviors: do fairness perceptions influence employee citizenship? J App Psychol. 1991;76(6):845–55.

58a. Theorell T. Commentary on Organisational Justice and Health of Employees: prospective cohort study. Occup Environ Med. 2003;60:33–4.

58b. Kivimaki M, Elovainio M, Vahtera J, Ferrie JE. Organisational justice and health of employees: prospective cohort study. Occup Environ Med. 2003;60:27–34.

59. Elovainio M, Ferrie JE, Gimeno D, De Vogli R, Shipley M, Brunner EJ, et al. Organizational justice and sleeping problems: the Whitehall II study. Psychosom Med. 2009;71(3):334–40.

60. Ferrie JE, Head J, Shipley MJ, Vahtera J, Marmot MG, Kivimäki M. Injustice at work and incidence of psychiatric morbidity: the Whitehall II study. Occup Environ Med. 2006;63(7):443–50.

61. Goldin C. From the valley to the summit: the quiet revolution that transformed women's work. National Bureau of Economic Research. 2004; No. w10335.

62. Bianchi SM, Milkie MA. Work and family research in the first decade of the 21st century. J Marriage Fam. 2010;72(3):705–25.

63. Moen P, Kaduk A, Kelly EL, Kossek E, Hammer L, Buxton OM, et al. Is work-family conflict a multi-level stressor linking job conditions to mental health? Evidence from the work Family and Health Network. Research in the Sociology of Work: Work & Family in the New Economy. Psychology. ;25.

64. Netemeyer RG, Boles JS, McMurrian R. Development and validation of work-family conflict and family-work conflict scales. J Appl Psychol. 1996;81(4):400–10.

65. Martikainen P. Women's employment, marriage, motherhood and mortality: a test of the multiple role and role accumulation hypotheses. Soc Sci Med. 1995;40(2):199–212.

66. Arber S, Gilbert GN, Dale A. Paid employment and women's health: a benefit or a source of role strain? Sociol Health Ill. 1985;7(3):375–400.

67. Frone MR, Yardley JK, Markel KS. Developing and testing an integrative model of the work–family interface. J Vocat Behav. 1997;50(2):145–67.

68. Chandola T, Kuper H, Singh-Manoux A, Bartley M, Marmot M. The effect of control at home on CHD events in the Whitehall II study: gender differences in psychosocial domestic pathways to social inequalities in CHD. Soc Sci Med. 2004;58(8):1501–9.

69. Wang J, Afifi TO, Cox B, Sareen J. Work–family conflict and mental disorders in the United States: Cross-sectional findings from the National Comorbidity Survey. Am J Ind Med. 2007;50(2):143–9.

70. Grant-Vallone EJ, Donaldson SI. Consequences of work-family conflict on employee well-being over time. Work Stress. 2001 Jul;15(3):214–26.

71. Frone MR, Barnes GM, Farrell MP. Relationship of work-family conflict to substance use among employed mothers: The role of negative affect. J Marriage Fam. National Council on Family Relations; 1994;56(4):1019–30.

72. Grzywacz JG, Marks NF. Family, work, work-family spillover, and problem drinking during midlife. J Marriage Fam. 2000;62:336–48.

73. Roos E, Lahelma E, Rahkonen O. Work-family conflicts and drinking behaviours among employed women and men. Drug Alcohol Depend. 2006;83(1):49–56.

74. Melchior M, Berkman LF, Niedhammer I, Zins M, Goldberg M. The mental health effects of multiple work and family demands: a prospective study of psychiatric sickness absence in the French GAZEL study. Soc Psychiatry Psychiatr Epidemiol. 2007;42(7):573–82.

75. Berkman LF, Buxton O, Ertel K, Okechukwu C. Managers' practices related to work–family balance predict employee cardiovascular risk and sleep duration in extended care settings. J Occup Health Psychol. 2010;15(3):316–29.

76. Geiger-Brown JM, Lee CJ, Trinkoff AM. The role of work schedules in occupational health and safety. In: Gatchel RJ, Schultz IZ, editors. *Handbook of occupational health and wellness.* Boston, MA: Springer; 2012. pp. 297–322.

77. Vyas MV, Garg AX, Iansavichus AV, Costella J, Donner A, Laugsand LE, et al. Shift work and vascular events: systematic review and meta-analysis. BMJ. 2012;345:e4800–0.

78. Jia Y, Lu Y, Wu K, Lin Q, Shen W, Zhu M, et al. Does night work increase the risk of breast cancer? A systematic review and meta-analysis of epidemiological studies. Cancer Epidemiology. 2013;37(3):197–206.

79. Akerstedt T. Shift work and disturbed sleep/wakefulness. Occup Med. 2003;53(2):89–94.

80. Costa G. Shift work and occupational medicine: an overview. Occup Med. 2003;53(2):83–8.

81. Sallinen M, Kecklund G. Shift work, sleep, and sleepiness: differences between shift schedules and systems. Scand J Work Env Hea. 2010;36(2):121–33.

82. Karlsson BH, Knutsson AK, Lindahl BO, Alfredsson LS. Metabolic disturbances in male workers with rotating three-shift work: results of the WOLF study. Int Arch Occ Env Hea. 2003;76(6):424–30.

83. van Mark A, Weiler SW, Schroder M, Otto A, Jauch-Chara K, Groneberg DA, et al. The impact of shift work induced chronic circadian disruption on IL-6 and TNF-alpha immune responses. J Occup Med Toxicol. 2010;5:18.

84. Knutsson A. Shift work and coronary heart disease. Scand J Soc Med Suppl. 1989 Jan 1;44:1–36.

85. Wang JL, Lesage A, Schmitz N, Drapeau A. The relationship between work stress and mental disorders in men and women: findings from a population-based study. J Epidemiol Community Health. 2008;62(1):42–7.

86. Straif K, Baan R, Grosse Y, Secretan B, Ghissassi El F, Bouvard V, et al. Carcinogenicity of shift-work, painting, and fire-fighting. The Lancet Oncology. 2007;8(12):1065–6.

87. Wang XS, Armstrong MEG, Cairns BJ, Key TJ, Travis RC. Shift work and chronic disease: the epidemiological evidence. Occup Med. 2011;61(2):78–89.

88. Kawachi I. Globalization and workers' health. Ind Health. 2008;46(5):421–3.

89. Sieber SD. Toward a theory of role accumulation. Am Sociol Rev. 39(4):567–78.

90. McNall LA, Masuda AD, Nicklin JM. Flexible Work arrangements, job satisfaction, and turnover intentions: the mediating role of work-to-family enrichment. J Psychol. 2009;144(1):61–81.

91. Greenhaus JH, Powell GN. When work and family are allies: a theory of work-family enrichment. Acad Manage Rev. 2006;31(1):72–92.
92. Kelly EL, Moen P, Tranby E. Changing workplaces to reduce work-family conflict: schedule control in a white-collar organization. Am Sociol Rev. 2011;76(2):265–90.
93. Galinsky E, Sakai K, Wigton T. Workplace flexibility: from research to action. Future Children. 2011;21(2):141–61.
94. Tausig M, Fenwick R. Unbinding time: alternate work schedules and work-life balance. J Fam Econ Issues. 2001;22(2):101–19.
95. Krieger N. Embodiment: a conceptual glossary for epidemiology. J Epidemiol Community Health. 2005;59(5):350–5.
96. Marmot MG, Bosma H, Hemingway H, Brunner E, Stansfeld S. Contribution of job control and other risk factors to social variations in coronary heart disease incidence. Lancet. 1997;350(9073):235–9.
97. Davey Smith G, Harding S. Is control at work the key to socioeconomic gradients in mortality? In: Davey Smith G, editor. *Health inequalities: Lifecourse approaches.* Bristol: The Policy Press; 2003. pp. 83–6.
98. Theorell T, Karasek RA. Current issues relating to psychosocial job strain and cardiovascular disease research. J Occup Health Psychol. 1996;1(1):9–26.
99. Siegrist J, Dragano N, Nyberg ST, Lunau T, Alfredsson L, Erbel R, et al. Validating abbreviated measures of effort-reward imbalance at work in European cohort studies: the IPD-Work consortium. Int Arch Occup Environ Health. 2014;87(3):249–56.
100. Work FHN. Work, Family, and Health Network [Internet]. projects.iq.harvard.edu. Cambridge, MA; [cited 2013 Dec 28]. Available from: http://projects.iq.harvard.edu/wfhn
101. NIOSH C. Total Worker Health [Internet]. cdc.gov. [cited 2013 Dec 28]. Available from: http://www.cdc.gov/niosh/twh/
102. Johnson JV, Hall EM. Job strain, work place social support, and cardiovascular disease: a cross-sectional study of a random sample of the Swedish working population. Am J Public Health. 1988;78(10):1336–42.
103. Fletcher BC, Jones F. A refutation of Karasek's demand–discretion model of occupational stress with a range of dependent measures. J Organ Behav. 1993;14(4):319–30.
104. Karasek R, Brisson C, Kawakami N, Houtman I, Bongers P, Amick B. The Job Content Questionnaire (JCQ): An instrument for internationally comparative assessments of psychosocial job characteristics. J Occup Health Psych. 1998;3(4):322–55. Special Section: The Measurement of Stress at Work.
105. Williams S, Cooper CL. Measuring occupational stress: development of the Pressure Management Indicator. J Occup Health Psych. 1998;3(4):306–21.
106. Siegrist J, Starke D, Chandola T, Godin I, Marmot M, Niedhammer I, Peter R. The measurement of effort–reward imbalance at work: European comparisons. Soc Sci Med. 2004;58(8):1483–99.
107. Kopelman RE, Greenhaus JH, Connolly TF. A model of work, family, and interrole conflict: a construct validation study. Organ Behav Hum Perf. 1983;32(2):198–215.
108. Pleck JH, Staines GL, Lang L. Conflicts between work and family. Monthly Labor Review. 1980;103(3):29–31.
109. Hackman JR, Oldham GR. Development of the Job Diagnostic Survey. J Appl Psychol. 1975;60(2):159–70.
110. Shinn M, Wong NW, Simko PA, Ortiz-Torres B. Promoting the well-being of working parents: coping, social support, and flexible job schedules. Am J Commun Psychol. 1989;17(1):31–55.
111. Eisenberger R, Cummings J, Armeli S, Lynch P. Perceived organizational support, discretionary treatment, and job satisfaction. J Appl Psychol. 1997;82(5):812–20.
112. Eisenberger R, Huntington R, Hutchinson S, Sowa D. Perceived organizational support. J Appl Psychol. 1983;71(3):500–7.

113. Carlson DS, Kacmar KM, Williams LJ. Construction and initial validation of a multidimensional measure of work–family conflict. J Vocat Behav. 2000;56(2):249–76.

114. Bronfenbrenner, U. Contexts of child rearing: Problems and prospects. Am Psychol. 1979;34(10):844.

115. Kossek EE, Colquitt JA, Noe RA. Caregiving decisions, well-being, and performance: the effects of place and provider as a function of dependent type and work-family climates. Acad Manage J. 2001;44(1):29–44.

116. Clark SC. Work cultures and work/family balance. J Vocat Behav. 2001;58(3):348–65.

117. Allen TD. Family-supportive work environments: the role of organizational perceptions. J Vocat Behav. 2001;58:414–35.

118. Carlson DS, Kacmar KM, Wayne JH, Grzywacz, JG. Measuring the positive side of the work–family interface: development and validation of a work–family enrichment scale. J Vocat Behav. 2006;68(1):131–64.

119. Hammer LB, Kossek E, Bodner T, Crain T. Measurement development and validation of the Family Supportive Supervisor Behavior Short-Form (FSSB-SF). J Occup Health Psych. 2013;18(3):285–296.

120. Eaton SC. If you can use them: flexibility policies, organizational commitment, and perceived performance. Ind Relat. 2003;42(2):145–67.

CHAPTER 6

LABOR MARKETS, EMPLOYMENT POLICIES, AND HEALTH

Mauricio Avendano and Lisa F. Berkman

INTRODUCTION

There is a positive association between employment and health: At any given point in time, those working are in better health than those unemployed or out of the labor force. This is extraordinarily consistent regardless of country, socioeconomic status, race, gender, age group, or marital status. It is tempting to conclude from this association that working is good for health, while unemployment or other forms of leave from work harm health. However, research over the last decades has revealed the complexity of establishing the intricate causal mechanisms generating a positive association between health and employment, which even to this day remain incomplete. On the one hand, a strand of literature suggests that working is good for health, and that negative "shocks" to employment, such as job loss, and perhaps retirement, lead to poor health. Conversely, work itself may harm health by exposing individuals to risky environments; work may also increase the opportunity costs of time, leading to lower investments in health than would otherwise be the case for an individual out of work. At the same time, it has now become clear that an important part of the relationship between employment and health reflects what social epidemiologists refer to as "reverse causality," or the impact of ill health on the ability to work.

The causal nature of this relationship has important policy implications: If working is good for health, policies that influence the length, continuity, and nature of employment are likely to impact health. For example, changes in employment law since the mid-1980s in many countries have led to a growing number of workers in fixed-term contracts, which may have had health consequences. Maternity leave policies, which extend women the right to take a period of leave from work around childbirth, are believed to influence female labor market attachment and career trajectories (1–5), and may through these mechanisms influence maternal and child health. Academics

and policymakers alike have intensely debated whether policies that promote longer working lives through later retirement might harm health or whether earlier retirement and shorter working lives may instead be health preserving. Both negative and positive spillovers to health may result from policies that shape these aspects of the dynamics of employment.

Key to embracing the relevance of these questions is the concept of human capital—broadly understood as the stock of knowledge, competencies, or other personal attributes that an individual has and that contribute to his or her "productivity"—and the idea of health as a form of human capital. Complementarities between health and other forms of human capital are also important. Job loss may lead to a loss of human capital by eroding skills and earnings, which may in turn lead to poor health in the long run. Conversely, protected job interruptions for childcare through paid maternity leave may help in keeping women attached to the labor market and reduce the human capital loss for women who would otherwise have to quit their jobs, potentially leading to long-run health gains for both mother and child. Similarly, policies that regulate retirement age may have consequences for the human capital investments workers make over their working life, and in turn, retirement itself may have direct implications for health and productivity in later life.

In this chapter, we examine the relationship between employment and health, with the aim to understand how policies that shape the dynamics of employment influence health and longevity. We critically examine evidence for the causal impact of employment and labor market transitions on health. We discuss an extensive literature on the impact of unemployment, job security, maternity leave, and retirement on health, particularly focusing on recent attempts to establish to what extent these associations are causal. We bring together this evidence to build a case for the significance for population health of policies that shape the dynamics of employment.

The chapter is organized as follows: We start by outlining a theoretical framework based on human capital theory. We argue that human capital theory may offer a useful framework to understand the link between work, employment policy, and health. This is followed by a critical review of theories and evidence on the impact of employment and job insecurity on health, with a particular focus on literature attempting to disentangle causality. We summarize a separate but related literature on the impact of business cycles on health, which goes beyond the direct effects of individual job loss, raising key questions about what are essentially unintended consequences of the economy on health. We then focus on how labor market transitions associated with maternity leave and retirement might impact health. We focus on these transitions, given their relevance for current societies experiencing major changes related to population aging and increasing conflict between work and family demands as women have joined the paid labor force. Policies that influence mothers' employment and the age of retirement may thus exert a potent force on the health and well-being of future populations. Based on an assessment of evidence, we draw implications for how policies that regulate the length, continuity, and nature of employment might impact health.

A THEORETICAL MODEL BASED ON HUMAN CAPITAL THEORY

A theoretical framework that outlines causal mechanisms is important because it identifies specific policies on which to intervene. Traditionally, many of the theoretical models motivating

unemployment research draw on psychological and social theories. These theories are useful as they propose several potential mechanisms through which labor market transitions influence health. However, they offer an incomplete framework of the more complex bidirectional mechanisms linking employment and health. Human capital theory (6–9) offers a potential unifying framework to understand the links between employment, labor policies, and health. Based on this framework, we argue that policies that promote a healthy and skilled population with strong labor market attachment—such as maternity leave policies—build up human capital and bring potentially long-run benefits to health. In contrast, employment policies that do not generate investments in human capital will produce a sick and unskilled workforce, which will in turn be less productive and erode human capital accumulation. This section outlines the basic principles of the Grossman model as a useful framework for understanding these relationships.

A SHORT SUMMARY OF THE GROSSMAN MODEL

The Grossman model is a *rational actor* theory of the demand for health that views each individual as both a producer and a consumer of health. The model is particularly useful in understanding the bidirectional relationship between health and labor market behavior (10, 11). Health is considered a stock determined by an individual's investments in health, and in this way it is conceptualized as a *capital*. Health is thus considered both a *consumption good* that provides utility, that is, individuals derive utility or "happiness" from good health, and an *investment good*, that is, health increases productivity, reduces time in illness, and yields higher wages by enabling individuals to work and earn income. According to this model, individuals invest time and other health inputs in order to produce health, and in this way maximize the time available for leisure and market activities. At the same time, investment in work is necessary to increase the income available to purchase material health inputs and other commodities from which individuals derive utility. This model therefore assumes that utility is a function of the stock of health of an individual, the consumption of other goods and leisure time. Each individual maximizes utility subject to a budget constraint, related to an individual's assets and the price associated with health inputs, and a time constraint. An important feature of this model in interpreting the links between work and health is that investments in health are costly: Individuals must trade off time and resources to produce health, such as time for cooking healthy foods and exercising, against time for working.

The Grossman model predicts several effects of health on labor market behavior: Health status determines the total time an individual spends in sickness, and therefore determines the total time available for market and nonmarket activities. A negative health shock (e.g., a new chronic disease diagnosis) may reduce the ability to work, thus diminishing productivity, which may translate into lower wages. In turn, lower wages may have both income and substitution effects, so that the net effect is unclear. Poor health may affect the marginal rate of substitution between leisure and health, because it increases the "disutility of work" (individuals are less able to derive utility when they are sick), reducing labor supply. At the same time, poor health may increase labor supply because ill individuals need income to acquire 'material health inputs' (such as healthy food, medical care and housing) (11).

According to this model, the labor market may also influence health through different channels. First, labor supply determines labor income, which determines the ability to purchase material health inputs. In this sense, more labor supply may be beneficial for health, for

example, by enabling workers to purchase healthier foods or exercise equipment, and provide better access to housing, healthcare, clothing, transportation, or other health-enhancing goods. On the other hand, labor supply decreases the time available to produce time health inputs (e.g., exercise, cooking healthy foods), and may therefore inadvertently lead to health deterioration. This second mechanism, often ignored by social epidemiologists, departs from the assumption that employment is always good for health. While employment yields health benefits, there is a trade-off between investments for health and investments in work that will need to be considered in understanding the relationship between work and health. Labor supply may also be a direct input in the *health production function*; for example, hazardous working conditions, job stress, and the physical exertion of employment may directly lead to poor physical and mental health.

Predictions from this model provide a general framework to understand why the association between labor force behavior and health is bidirectional and, in some sense, contradictory: Working increases income to purchase material goods that are important to health, but at the same time, working reduces time to invest in health, or may directly harm health through exposure to hazardous working conditions among lower-wage workers. These conflicting mechanisms will become more evident when discussing the discrepancies between studies on the impact of unemployment on health at the individual level, and studies on the aggregate business cycles and population health.

HUMAN CAPITAL AND EMPLOYMENT PROTECTION POLICIES

Social protection can be conceptualized as a way to help economic actors to overcome market failures in human capital formation. By addressing these market failures, social protection policies may promote accumulation of health capital. Maternity leave policies might reduce gender differences in labor market trajectories, labor supply, and accumulation of human capital due to fertility decisions and childcare (1, 12). They may increase labor market attachment by making it more likely for women to return to their employer (1), hence reducing losses in human capital associated with extended periods out of the labor force. In a model of enduring employment relations with no maternity leave benefits (12), a new mother will be obliged to quit her job and lose firm-specific human capital, bearing the costs of searching for a new job. The introduction of maternity leave reduces the cost of job interruptions for the mother; it increases job protection and female labor attachment by increasing the time employed (even if temporarily not at work) and reducing time spent out of the labor market (1, 4, 12). Maternity leave policies may thus increase job continuity and help mothers retain use of skills or knowledge specific to their previous employer. This may enhance productivity and result in better long-term earnings and career advancement, potentially leading to long-term gains in health (4).

Maternity leave policies may also yield important human capital benefits for children. A growing literature recognizes that parental employment plays a powerful role in determining the resources families devote to children (4, 13). Maternity leave policies enable the mother to spend time with her child during the critical years of early development, leading to potential gains in human capital accumulation, for example, by yielding better cognitive and educational outcomes.

Maternity leave benefits may thus bring both short- as well as long-run benefits to child health (4, 14–16).

Policies on retirement are also strongly intertwined with human capital accumulation. Most Western countries will be faced with the consequences of demographic aging in the coming decades, with most workers living longer after retirement. Economic theory predicts that workers will invest more in skills and human capital over their working life if they expect to retire later, as compared with a scenario where they expect to retire early. Put another way, individuals have a stronger incentive to retire later if they have acquired more human capital on the job, since this raises forgone labor earnings while being retired. This rationale is often used as an argument to raise statutory retirement ages and promote later retirement. While there is much debate on the empirical basis for these theoretical claims, a key message is that retirement policies influence human capital investment decisions over the working life, such as acquiring on-the-job training, which in turn influence career and earning trajectories over the lifecourse. Through this mechanism, retirement policies may have potentially important consequences for physical and cognitive health in the post-retirement years.

EMPLOYMENT AND HEALTH: UNDERSTANDING THE RELATIONSHIP

The impact of employment—or the lack of it—on health has been the focus of much research in epidemiology, sociology, and psychology. Janlert and Hammarström (17) identified three historical traditions concerning the relationship between employment and health: the *biomedical tradition*, which focuses on the physiological and biological mechanisms explaining the relationship between physical health and employment; the *sociological tradition*, which focuses on the material circumstances generated by the lack of employment and their impact on health; and the *psychological tradition*, which focuses on the psychological effects of unemployment and their impact on health. In a separate review, Bartley summarized the theories to explain the association between unemployment on health into three broad causation models (18). In this section, we integrate these two complementary classifications. While this literature has been developed in the context of research on unemployment, we argue that it offers general insights on how employment more generally relates to health. These theories are therefore also relevant for understanding the health effects of other labor market transitions, particularly maternity leave and retirement. There are several potential models of how employment may influence health:

ECONOMIC DEPRIVATION MODELS

This perspective, which represents the classical approach in sociology, argues that changes in family income induced by unemployment hamper the ability to accumulate wealth, and reduce access to material resources that are relevant to health. In support of this hypothesis, studies suggest that financial strain and uncertainty is a strong mediating factor between unemployment and ill health (19, 20). This is also supported by studies comparing the health impact of unemployment among

workers who expected to receive generous early retirement benefits to that of workers who were not eligible for such benefits (21).

A policy implication from this theory is that income benefits during unemployment, maternity leave, or retirement should mitigate any negative impact of these labor market transitions on health (17). Unemployment cash benefits should release financial pressures and smooth consumption, thus preventing health deterioration after job loss. Legislation that extends women the right to *paid maternity leave* should mitigate the financial stress associated with a period out of work due to maternity, not only during the period around birth but also in the long run, by ensuring that women can return to their previous employer and continue their career trajectories (1–5). Income benefits after retirement should smooth consumption across the life cycle and prevent financial strain at old age, which should mitigate the negative effects of retirement.

While useful, it is by now clear that the economic deprivation model offers a narrow view of how labor market transitions influence health. A link between unemployment and health has been observed even in countries with generous unemployment benefit programs, such as Finland (22, 23) and Sweden (24, 25). From a policy standpoint, income benefits may smooth consumption and reduce financial strain during periods out of the labor market, but the mechanisms through which income benefits influence health appear to be more complex. For example, extending the duration of unemployment insurance benefits may inadvertently increase the duration of nonemployment spells (26); likewise, extending maternity leave benefits may delay mothers' return to work, although not necessarily jeopardizing labor market attachment (27). Inadvertently, these two somewhat undesirable consequences of income transfer policies may have negative health effects. On the other hand, income protection may improve health by reducing the human capital loss of job interruptions: Extended income benefits may enable displaced workers to find a job that better matches their skills and levels, or they may allow women with children to remain attached to their previous employer, improving their long-term labor market outcomes, and ultimately influencing health. Overall, a model that narrowly centers around the direct effects of economic deprivation on health during employment gaps does not capture the potentially more important long-run effects of income support during periods out of the labor market on human capital accumulation and health in the long run.

THE NONFINANCIAL BENEFITS OF WORK

First proposed by Marie Jahoda (28), the theory of latent functions argues that, in modern welfare states, the threat of starvation and physical deprivation no longer accompanies the loss of paid work. This implies that work provides not only manifest (e.g., wages) but also latent functions. Unemployment is understood as a threat to health because work has a number of nonfinancial benefits, or latent functions, such as providing a time structure for the day, opportunities for social contact with other people, self-esteem and status formation, and a sense of contributing to a collective purpose. Warr (29) expanded this theory on the benefits of work for mental health, including other latent functions such as physical and mental activity, use of skills, decision latitude, interpersonal contact, social status, and "traction"—a motivation to go on from one day to the next (18).

A policy implication from this theory is that, contrary to the economic deprivation model, income benefit programs may be insufficient in mitigating the impact of employment gaps on

health. This view is more in line with a model that conceives of employment as an opportunity to build up human and social capital. For example, this model would predict that policies that promote later retirement or incentivize retirees to engage in meaningful productive activities after retirement may bring health benefits. Similarly, policies that encourage human capital investments during unemployment spells, such as active labor market programs, might partly compensate for the human capital loss associated with unemployment. Maternity leave policies may also enable women to engage in childcare in a more productive way, partly compensating for the temporary loss of benefits from employment during maternity leave. As with the economic deprivation model, however, this theory focuses on the short-term benefits of employment, but it does not consider the long-run cumulative effects of employment on careers, earnings, and other social determinants of health in the long run.

THE STRESS MODEL

This perspective, which originates from stress models in psychology, conceptualizes unemployment as a psychosocial stimulus that triggers a stress mechanism, acting as precursor for the development of disease. An important concept in this model is "coping," which refers to the ability of the individual to handle the stress generated by unemployment. It conceptualizes stress as a chronic process, which influences physical health as a result of chronically increased levels of anxiety. It suggests that unemployment is a stressor potentially leading to heightened arousal, distress, withdrawal, and lower motivation, which may lead to chronic disease (30). This perspective is congruent with the view that unemployment triggers chronic stress-related pathways, which over the long term have a deleterious cumulative effect on health (18, 31).

A related concept within this model refers to the role of control, which refers to the possibility that unemployment leads to a lack of control on the environment (32, 33). This approach assumes that individuals want to have primary control over their own environments. Failure to exert primary control may produce frustration, loss of self-esteem, and loss of confidence, leading to helplessness and depression (33). The magnitude of this effect depends on the ability of the individual to adapt to the new situation. A common model in relation to this hypothesis derives from the demand-control theories (34). These models, discussed extensively in Chapter 5, conceive of unemployment as a passive work situation, with low demands and low control in relation to working (17). These models provide an overarching psychological framework for understanding the mental health consequences of unemployment (32).

An implication of this model is that policies that ameliorate stress associated with employment gaps may be important in mitigating any effects on health. To the extent that stress is the result of financial strain, income benefit programs may be able to partly buffer the impact of periods out of the labor force on health. On the other hand, stress may also result from anxiety about the long-term prospects of returning to the labor market after job loss or a period of inactivity associated with maternity leave. In this scenario, employment protection policies that diminish job insecurity might reduce stress and bring benefits for the health of the employed, although they may offer little benefit to those out of the labor market. It is difficult to envisage other specific policies stemming from this model other than those that directly address the stress effects of unemployment, such as stress management and counseling programs for the unemployed. While potentially

a buffer, such programs are unlikely to completely reverse the long-run effects of employment gaps on career trajectories, earnings, and human capital accumulation, and their effects on health in the long run.

SOCIAL SUPPORT MODEL

As discussed in Chapter 7 of this book, the extent of social support and integration has long been linked to health. Unemployment, maternity leave, or retirement may lead to a loss of social support and networks at work, which may negatively influence health. In addition, unemployment or diminished household income during maternity leave may strain family relations. Social support may also act as a safeguard, with individuals with higher levels of support suffering fewer health consequences from unemployment, maternity leave, or retirement. On the other hand, periods of unemployment or maternity leave may also decrease the opportunity costs of time and lead to more social participation, which may ultimately lead to better health.

HEALTH-RELATED BEHAVIOR

Evidence suggests that individuals who are unemployed have increased risks of some forms of health-damaging behavior, particularly heavier smoking, drinking, and leisure physical activity (18). There are two reasons for this association: First, individuals who smoke or drink heavily may be more likely to become unemployed than their healthier counterparts. Second, individuals may change their behavior as a result of unemployment. The latter can involve two mechanisms in different directions: Unemployment may reduce social interactions, leading to less, light, or moderate alcohol consumption. However, unemployment may also increase smoking and alcohol consumption, as individuals become more isolated and consequently may use these behaviors as a way to deal with unemployment.

Whether behavior changes develop before unemployment or whether they are an outcome of job loss remains a matter of debate. Some evidence from longitudinal studies suggests that as a result of increased financial strain and reduced social interaction, the amount of smoking and alcohol consumption might actually decrease during unemployment spells (35). This pattern is consistent with reductions in aggregate levels of smoking during economic downturns observed by Ruhm (36, 37). On the other hand, most studies find that the unemployed are more likely to smoke and drink heavily than their employed counterparts. However, there is limited evidence that consumption of these goods increases after unemployment, potentially reflecting selection into unemployment among those with less healthy behaviors.

A policy implication from this model is that in order to reduce the effects of unemployment or retirement on health, interventions that promote healthy behavior are most important. However, given limited evidence of a causal effect of unemployment on behavior, it is yet uncertain that behavioral interventions during unemployment or retirement would lead to improvements on health. In addition, a model that narrowly emphasizes proximal behavioral determinants may be insufficient to account for the potentially complex processes that might incentivize unhealthy consumption during employment gaps or after retirement.

A LONG-RUN PERSPECTIVE ON EMPLOYMENT AND HEALTH

We have considered a series of possible models that provide a partial explanation for the association between employment and health. A common denominator to these theoretical models is the focus on relatively short-term mechanisms, and the idea that a "shock" that disturbs employment will lead to financial, material, social, psychological, or behavioral changes that will harm health. These theories, however, place very little emphasis on the potential long-run effects of employment on human capital accumulation, and the more complex ways through which employment shapes the lives of individuals and their long-run health.

Early in 1982, David T. Ellwood (38) used the term "scar" to describe the persistent, long-term effects of unemployment on wages and labor market outcomes (38). During unemployment, work experience cannot be accumulated, which leads to the erosion of skills, and prevents the acquisition of new skills. Unemployment increases the chances of repeat job interruptions leading to low-paid, unstable jobs and losses in current and future income. Unemployment may thus leave a permanent "scar" on a worker's career. This was first documented by Ruhm (39), who showed that among workers displaced during the mid-1970s, at the time of reemployment four to five years later their earnings were 10% to 13% lower than those of identical workers who were continuously employed (39). Effects of similar magnitude have been reported in the United Kingdom (40). These findings have been replicated by others showing that job loss is accompanied by a long period out of work, a major loss of earnings, and reduced career prospects for the next years or decades (40–43).

Periods out of the labor market due to maternity leave are likely to affect women's careers in a different way than unemployment, as they may be perceived less negatively by prospective employers. However, maternity leave, too, has been shown to scar women's careers, affecting their subsequent wage growth rate and career prospects (1, 5). Central to the idea of scarring is that even after re-employment, workers who temporarily leave the labor market may endure a long-term scar in their careers. An explanation often offered for this pattern is that at the time of engagement with new workers, prospective employers may use a worker's career history as a signal of productivity, penalizing those who have had a gap in their employment histories (40). As a result, recent mothers and displaced workers returning to work will be more likely to accept lower wages and less favorable contractual conditions than their more stable counterparts. This process drives wages down and erodes working conditions for workers returning to work, who will increasingly face lower wages and heightened job insecurity (39–42). These effects may have long-lasting consequences for workers' career, earnings, and wealth, and in turn result in poorer health prospects.

There are two important implications related to this theory that deserve attention. First, the effects of employment gaps on health may be *cumulative*, leading to a pathway of social and economic disadvantage that translates into poorer health outcomes many years later. This implies also the possibility of a long *etiologic period* between employment and health, with long-lasting health consequences of employment gaps only becoming evident in older ages. The second implication is that if job loss and maternity leave have scarring effects, re-employment will not enable complete health recovery, with some or all of the effect remaining after an employment gap has ended.

In this chapter, we emphasize the importance of a theoretical framework that shifts focus from short-term fluctuations in health in response to employment shocks, to a life cycle model that

emphasizes the cumulative nature of employment trajectories and human capital accumulation, and their impact on health in the long run. We argue that these long-run processes are more essential to understanding health and survival, especially as individuals approach old age and the cumulative impact of employment experiences over the lifecourse take their toll.

UNEMPLOYMENT AND HEALTH

In this section, we present a critical review of some of the literature addressing the causal relationship between unemployment and health. Research on this question appears to be strongly "countercyclical": The number of studies on unemployment and health decreases when the economy is expanding, and increases when the economy contracts (32, 44). This partly explains the re-emergence of interest on the health effects of unemployment during the last few years following the recent economic downturn that started in 2008. Over the last years, several reviews have summarized the evidence on this question (18, 45–47). Our aim is not to replicate these reviews, but instead to carry out a critical review of studies that focus on disentangling to what extent unemployment has a causal impact on health.

Earlier studies were based to a large extent on cross-sectional data, while more recent studies have primarily used longitudinal data and focused on involuntary job loss or job displacement. Involuntary job loss is often defined as discharge from paid employment for any reason when a worker would have preferred to continue working. This includes job loss due to factory closing, relocation, or downsizing; layoff or firing; the ending of a temporary job; or any other job loss when the worker did not voluntarily decide to terminate employment (48). Job displacement is a form of job loss that occurs when firms downsize, restructure, close factories or relocate, or when a worker is not recalled from a layoff. It is a specific type of job loss resulting from changes in the economy. As we shall see, the distinction between "exogenous" job loss—presumably independent of a worker's performance or previous health—and "endogenous" job loss—the result of poorer health or predictors of health—will become essential to the design of empirical studies testing whether the relationship between unemployment and health is causal.

Most studies linking unemployment and health conducted during the last decades are longitudinal, but approaches to ascertain whether associations are causal differ. Studies can also be classified according to whether they focus on mental health, physical health, or mortality outcomes. We primarily focus on distinguishing studies based on their methodological approach, but we discuss how different study designs have contributed to the understanding of the impact of unemployment on different health outcomes.

LONGITUDINAL STUDIES

In attempting to disentangle the impact of job loss on health, investigators typically compare the health of workers who experience job loss to a group of otherwise comparable workers who are continually employed. To rule out the possibility of selection and confounding, investigators control for a wide set of potential confounders, including baseline health and early life conditions.

Typically for the United States and the United Kingdom, studies are based on large cohort or panel studies. For the Scandinavian countries, and increasingly for other countries as well, studies are often based on linked-registry data that contain information routinely collected by government agencies on employment, education, demographic characteristics, health, and mortality. Panel surveys have the advantage of recording detailed information on a wide set of potential confounders and mechanisms involved. Registry-linked studies, on the other hand, have the advantage of incorporating entire national populations or large sample sizes. They do not suffer bias inherent to surveys based on self-report, and their large samples enable examining the impact of unemployment on relatively rare outcomes such as mortality. The main disadvantage of administrative data compared with survey data is the lack of detailed information on potential confounders, including earlier health status and skill sets and other variables containing information on the nature of unemployment spells. In addition, milder outcomes such as undiagnosed depression, and unemployment spells, are not recorded by administrative data if individuals do not attend medical services or register to claim unemployment benefits (49).

A first group of studies of this type relates health outcomes to earlier experiences of unemployment, whether due to job loss or other reasons. Invariably, the majority of studies in this category find that there is a strong association between unemployment and health that is robust when controlling for early life factors such as educational level, baseline income, preexisting health and other characteristics potentially correlated with both employment and health. In a classical study based on the British Household Panel (BHPS), a longitudinal survey of about 5,500 private households and 10,000 men and women interviewed every year in the period 1991–2001, Bartley and colleagues (50) showed that unemployment was associated with twice the hazard of limiting illness in the following year compared with continuous employment. As is typical in these studies, those not economically active (that is, those out of the labor force not actively looking for employment) were at even greater risk of illness compared to the employed. In a follow-up study based on the same data, Booker and Sacker (51) show that the onset of unemployment is associated with lower scores of psychological well-being as measured by the General Health Questionnaire (GHQ-12), with the first and second unemployment spells having a stronger effect than subsequent spells (51). These are examples of a large set of studies following a similar approach and confirming an association between unemployment and health across different countries and settings.

While these studies have several strengths and control for many potential confounders, the possibility that associations are due to selection or unmeasured confounding cannot be completely ruled out. For example, it is possible that workers who become unemployed are less healthy than those who continue to work and have therefore worse health prospects after unemployment, a difference that may not be completely captured by baseline adjustment. Individuals who become unemployed may also be different from those continually employed in a number of characteristics that are not observed in the data, such as parental characteristics, intellectual ability, time preference, and effort.

To overcome these limitations, recent studies attempt to distinguish experiences of unemployment that may be unrelated to an individual's former health or characteristics from those that may be health-related. A powerful advantage of these studies is that they distinguish involuntary job loss from potentially "voluntary" or health-related termination of a job contract. The most convincing studies are those looking at the effects of job displacement. Risk of disease among workers who experience job displacement presumably unrelated to personal characteristics is compared

with risk among continually employed workers. Assuming job displacement is "exogenous," an association can be interpreted as reflecting the causal impact of job loss on health, rather than the selection effects of preexisting health or other personal characteristics on the decision or ability to work. These studies are often labeled "natural experiments," because they exploit the natural "randomness" of plant closures or business cycle fluctuations in order to identify the causal effect of job loss on health.

In an important series of studies, Gallo and colleagues employed data from the Health and Retirement Survey (HRS) to examine how job loss in the years leading up to retirement affects physical and mental health outcomes. They used data from the HRS (52) to examine whether involuntary job loss, defined as job loss due to either a business/plant closing or layoff, was associated with subsequent risk of myocardial infarction and stroke. They compared the risk of these outcomes in 582 workers who experienced involuntary job loss over a 10-year follow-up with risks among 3,719 individuals who continued to be employed over the follow-up period. Controlling for well-established risk factors for these outcomes, results indicated that displaced workers had a more than twofold increase in the risk of subsequent myocardial infarction and stroke relative to employed individuals (52, 53). Following a similar approach, they documented a similar association between late-career involuntary job loss and physical functioning on mental health (54); the onset of alcohol drinking among nondrinkers (55); smoking relapse and intensity among displaced workers with a history of smoking (56); and depressive symptoms among displaced workers with limited wealth (57) and higher education (58). A limitation to keep in mind is that the reason for job loss in these studies is based on respondents' own reports, rather than on an objective assessment of firm decisions in relation to workers. In addition, some workers may anticipate firm closure or mass lay-offs and move to other companies before having to involuntarily leave their jobs. Notwithstanding these limitations, these studies suggest that as workers approach retirement, late-career job loss may have extensive physical and mental health consequences.

Another example of this approach comes from a study based on two large population-based longitudinal samples of US workers from the US Changing Lives Study (CLS) and the Wisconsin Longitudinal Study (WLS). For WLS respondents, Burgard and colleagues (48) were able to distinguish the reason and timing of events, distinguishing involuntary job loss due to plant closing, downsizing, or relocation; other involuntary termination due to firing or layoff; temporary or seasonal layoff; health-related reasons; or imprisonment. They examined the impact of different forms of job loss on self-rated health and depressive symptoms. Their results suggest that, controlling for an extensive set of confounders, involuntary job loss is associated with subsequent poorer self-rated health and more depressive symptoms. However, effects on self-rated health were primarily driven by health deterioration among those who were initially in poor health before becoming unemployed, suggesting that job loss may be more of an amplifier of an existing health problem. Nevertheless, there was a small but significant effect of job displacement on self-rated health. By contrast, job loss had a strong and consistent effect on depressive symptoms regardless of preexisting health (48).

In a separate study, Strully (59) used data from the Prospective Study of Income Dynamics (PSID) to assess the impact of job loss on health. The author distinguished job separations due to a workplace closure, a form of job loss less vulnerable to selection bias, from job loss due to firing or layoff, voluntary job separation, or other forms of job separations more prone to selection bias. The study linked these experiences of job loss to subsequent reports of poor self-rated health;

the onset of health conditions likely to arise in the short-term after job loss, such as stroke, heart disease, or psychiatric conditions; and the onset of a long-term health condition unlikely to arise in the short-term after job loss, such as lung cancer or memory decline. Findings suggest that, controlling for prior health status, job loss due to workplace closure increases the odds of poor health by 54% and the odds of a new likely health condition by 83%. These results were not explained by health-related section. Although other forms of job loss were also related to health, controlling for baseline health strongly attenuated these effects, suggesting a strong role for health selection in job losses unrelated to closures.

The strength of these and similar studies is the distinction between involuntary job loss and other forms of unemployment potentially more related to individual decisions or personal characteristics. These studies provide some evidence of a detrimental effect of unemployment on health that is not due to selection. Although information on exposure comes from self-report, these studies are considered natural experiments because they exploit episodes of job loss due to reasons that fall presumably out of the control of individuals, such as firm closures or mass layoffs. On the other hand, involuntary job loss as defined in these studies does not completely preclude selection. Particularly job loss due to layoff is unlikely to result from random processes affecting all workers in a given firm. For example, individuals with mental health or physical health problems, or those with fewer skills or commitment, may be more likely to be laid off than their healthier and more committed and skilled counterparts. Job loss due to firm closure, on the other hand, is more likely to be "exogenous" or unrelated to personal characteristics, although selection cannot be completely ruled out. For example, workers may be sorted into firms with different likelihood of future closure based on their individual characteristics, that is, successful firms may be more able to recruit healthier and more skilled workers than firms with a higher risk of default. Nevertheless, these studies are a large step forward and suggest that at least part of the association between unemployment and health may be due to the detrimental effects of job loss on health.

ECONOMETRIC STUDIES

Despite efforts at controlling for selection by focusing on job loss due to business closure, a concern in studies cited is the possibility that "treated" workers—those experiencing job loss—are different from "control" workers—those continuing to work, in unmeasured variables relevant to health. To address this possibility, econometric studies have started to apply more sophisticated techniques including difference-in-differences approaches, fixed effect models, and propensity score matching. In this section, we review some of the most recent studies using these approaches.

Using data from the HRS, Salm (60) estimated the effect of job loss on health for workers nearing retirement. Unlike the series of HRS papers by Gallo and colleagues (52–56, 61, 62), he more directly addressed reverse causality by focusing on workers laid off due to closure of their previous employer's business, using a difference-in-differences approach. The author compared health changes across waves in a sample of workers who experienced job loss due to their employer's business closures, with health changes in a sample of workers who did not experience job loss. Surprisingly, he found no effect of job loss due to business closure on various measures of physical and mental health, even among population subgroups expected to be more affected by job loss. These findings contrast with the findings by Gallo and colleagues (52–56, 61, 62),

who found that job loss due to plant closure or lay-off was associated with several physical and mental health outcomes. Using a similar difference-in-differences approach supplemented by propensity score matching, Bockerman (63) found that unemployment does not influence self-rated health in Finland, and concluded that the association is likely due to health-related selection into unemployment.

In a separate study, Browning and colleagues (64) used propensity score matching to investigate whether job displacement causes hospitalization for stress-related diseases in Denmark. A random 10% sample of the Danish male population for the years 1981–1999, including information on demographics, health, and work status was individually linked to data with each worker's employer. Based on matching estimates, their results indicated that displacement in Denmark does not cause hospitalization, a finding that holds across multiple population subgroups. Based on an individual fixed effect model, Schmitz (65) used the German Socio-Economic Panel to assess how job loss due to business closure relates to short-term changes in health in Germany. Schmitz used an individual fixed effect model to control for unobserved heterogeneity across individuals. In a fixed effect model, variation across individuals is controlled for, so that estimation relies only on variation within individuals over time. This is equivalent to comparing health at a time of job loss for an individual at time *t* to health for the same individual at a later or earlier point when the individual was employed. An earlier study (66) based on the same approach and data, but not distinguishing different forms of job separation, found an effect on unemployment on health. In contrast, Schmitz (65) found that job loss due to business closure had no effect on health satisfaction, the probability of a hospital visit, or mental health scores. Job separation due to other reasons potentially related to health, on the other hand, was strongly associated with health changes, suggesting again that selection effects might drive the association between unemployment and health.

The studies described above are able to better isolate the causal effects of job loss on health than earlier studies that assumed treated and control groups to be exchangeable. Overall, these econometric studies find little evidence that job loss due to business closure has an impact on physical and mental health. There are several potential reasons for the discrepancy in findings between these studies and those from earlier investigations. First, it is possible that by using propensity score matching and difference-in-differences estimators, these studies are indeed better able to control for reverse causation and confounding. While we cannot discard this possibility, a potential limitation of these studies is that, although they often find no significant effects, their standard errors are nevertheless very large. The efficiency costs in fixed effect or differencing models is well known (67); these studies exploit within-individual variation in exposure to identify the effect, blocking all variation across individuals. Similarly, propensity score matching estimators rely on a small subset of all observations that could be matched, and therefore often result in huge standard errors. It may thus well be that these studies are simply underpowered to detect an effect of job loss on health, as most of them would only be based on a few hundred cases experiencing job loss due to business closure.

A second reason for the discrepancy in findings may stem from the fact that econometric studies restrict the definition of exposure to job loss due to business closure, while some of the earlier studies combine job loss due to business closure and lay-off, the latter being potentially the result of baseline health. However, two of the earlier studies (48, 59), which also distinguished job loss due to business closure, did find a small yet significant effect on health, controlling for baseline

health. A third explanation is that many of the econometric studies are less careful in defining their health outcomes as compared with earlier epidemiological studies. Many of these studies are based on overall self-assessments of health or health satisfaction. A possibility remains that job loss influences the risk of some outcomes but not others. For example, job loss may increase the risk of stroke and myocardial infarction (52) while having a weaker impact on measures of general self-assessments of health or health satisfaction.

An important issue to consider is that several of the studies that reported an effect of job loss on health are based on US data sets, such as the HRS, the CLS, or the WLS. In contrast, many of the recent studies examining the links between unemployment and health are based on data for European countries, including Finland, Denmark, and Germany, which have substantially stronger social and employment protection systems. A potential hypothesis is that job loss has worse health consequences for US workers enjoying limited social protection as compared with workers in many European countries, where levels of social protection are higher. A recent study, for example, found that the association between unemployment and mortality is much stronger in the United States than in Germany (68). However, this study did not distinguish "exogenous" job loss from health-related unemployment spells, and it is likely to reflect the bidirectional relationship between health and employment. The current literature offers a limited understanding of the potential institutional mechanisms that may modify the impact of job loss on health.

A final explanation is that the econometric studies above use techniques that only identify the short-term impact of job loss on health with no or a very short time lag, while some of the earlier studies follow participants over longer periods or until the onset of a given condition or health outcome (52–56, 61, 62). It is questionable whether fixed effect methods are an appropriate approach in the context of long-run, permanent effects that remain even after individuals experience re-employment. If the effects of job loss on health have long-run effects rather than responding to short-term fluctuations with no or short time lags, fixed effect models will be misspecified and yield biased estimates (69). Consideration of long-run effects is important because, as we shall see in the next section, job loss may have long-run effects on human capital accumulation with health consequences becoming evident only many years or decades after exposure.

SUMMARY

The evidence discussed above—which is by no means exhaustive—illustrates that there is not full agreement across an extensive number of studies using different methodologies. So far, the most convincing studies are natural experiments that exploit business closures or mass layoffs to identify the effects of job loss. The conclusions from these studies depend on the approaches used to control for selection and confounding; the specific outcomes examined; and the populations under study. Importantly, several studies using advanced econometric techniques find weak evidence of an effect of job loss on health. There is reason, however, to be skeptical, given the focus on short-term effects in these econometric designs; the large standard errors in estimates from fixed effects, difference-in-differences, and propensity score matching studies; and the lack of careful definitions of health outcomes and appropriate assessment of etiologic periods or lag structures.

In addition, many of these studies come from European countries with strong social protection systems, where the impact of job loss on health may be weaker than in the United States.

Notwithstanding the limitations of existing research and the challenges posed by econometric studies, there seems to be some supportive evidence for the hypothesis that job loss leads to poorer mental health. This seems to be most consistent for self-assessments of non-clinically diagnosed mental health, while findings for hospitalizations for serious mental illness appear less consistent. In addition, some studies suggest that job loss is associated with increases in some—but not all—physical health outcomes, including major cardiovascular diseases and other major illnesses. These effects, however, are likely to be much smaller than initially reported by earlier studies, as an important part of the association appears to be driven by less healthy individuals being more likely to become unemployed. Nevertheless, the fact seems to remain that job loss has a small yet significant effect on a variety of physical and mental health outcomes.

Part of the puzzle may be due to the fact that the studies above do not capture the long-run effects of job loss on health, but only capture short-term fluctuations of health during periods of unemployment. As such, it is not surprising that mental health outcomes are more sensitive to these short-term fluctuations. In addition, the lack of effects in some studies may also reflect features of the institutional context that may determine the extent to which job loss influences health. Unemployment benefits may diminish the long-term career effects of job loss by enabling workers to find a better match for their skills, buffering any career effects. In the short run, benefits also smooth consumption during unemployment spells, preventing health effects associated with income constraints. It is difficult to establish to what extent these and other institutional features may explain the lack of effects in some European studies. In addition, at least one econometric study in the United States also found no consistent effect of job loss on health (60). Nonetheless, these findings offer a new potential avenue for research to establish the institutional features that may shape the mechanisms linking unemployment to health.

THE LONG-RUN EFFECTS OF UNEMPLOYMENT ON MORTALITY

The idea that unemployment increases mortality has long been a topic of research, which some believed settled in 1984, after a study linking a 1% sample of the population of England and Wales census from 1971 to subsequent mortality in the period 1971–1981 showed a strong association between unemployment and mortality (70–72). Since then, much research has been conducted to expand this work and address the critical concerns of confounding and reverse causation. As in the earlier sections, our focus here is on recent studies trying to address the question of whether the association between individual job loss and mortality is causal.

In a recent meta-analysis, Roelfs and colleagues (73) concluded that unemployment is associated with on average a 63% increased risk of all-cause mortality, with associations being similar across the United States and European countries. This review, however, did not distinguish studies that addressed reverse causation from studies examining associations, so it is difficult to draw firm conclusions from this summary estimate. Recent evidence, however, would seem to suggest that unemployment may have stronger effects in some countries than in others. As mentioned earlier, in a recent study, Mcload and colleagues (68) documented an association between unemployment

and mortality in the United States, but no association was found in Germany. Again, however, these studies are prone to confounding and reverse causality.

An important turn in the literature, typically finding a strong association between unemployment and mortality (23, 74–78), came through studies that linked the state of the national economy to individual-level transitions into unemployment. In two important studies, Martikainen and colleagues (22, 79) examine how the association between unemployment and mortality changes during periods of high versus low aggregate unemployment in Finland. Their rationale was that in the context of strong selection, the association between unemployment and mortality should be stronger in a period of low unemployment than in a period of high unemployment, when both healthy and less healthy individuals would become susceptible to unemployment, thus leading to less selective job loss. In their first study (79), they found that although there was an association in both periods, unemployment was more weakly related to mortality in periods of high unemployment, when less health selection is likely to operate. Recent studies have found similar patterns for specific outcomes such as suicide (80). This suggests that at least part of the association between unemployment and mortality is due to selection.

In a second study, Martikainen and colleagues (22) examined whether job loss due to severe downsizing at a worker's firm was associated with increased mortality over the next four years. Again, they were interested in assessing this question separately for workers losing their job in a period of low unemployment (1989) compared with a period of high unemployment (1994). For this purpose, they used detailed individual-level data on unemployment spells from the National Office of Statistics for a representative sample of Finnish workers. They linked each worker to firm-level data from the Finland Establishment Register that included data on turnover, production, industry, and staffing levels for each firm. They then individually linked records to registry data on mortality from the national statistics office. Again, their study suggests that selection plays a large role in explaining the association between unemployment and mortality in Finland. Unemployment was associated with a more than twofold increase in the hazard of mortality after 1989 (the low unemployment period), compared to a 25% increased risk after 1994 (the high unemployment period). There was no evidence of increased mortality among workers who lost their job at firms that had experienced large reductions in size, and the association between unemployment and mortality was weaker among those working in firms that had been strongly downsized. The fact that the association between unemployment and mortality is weaker in periods of high unemployment suggests that while there may be an effect, this may be much smaller than previously thought, as selection into unemployment may explain a substantial part of the association.

Recent evidence for the United States, on the other hand, suggests that there may be a causal effect of job loss on mortality even after controlling for potential selection. In a recent study, Sullivan and Von Wachter (81) used administrative data on the quarterly unemployment of Pennsylvanian workers in the 1970s and 1980s linked to Social Security Administration death records covering 1980–2006 to estimate the effects of job displacement on mortality. They found that high-seniority male workers who had stable careers and lost their jobs as a consequence of aggressive downsizing during the major economic downturns in the 1970s and 1980s in Pennsylvania had a 50–100% increased risk of death in the year after displacement. Although the effect of job loss declines over time, even 20 years after displacement displaced workers had a 10–15% increased risk in annual death rates. Their results are robust to a number of strategies to address potential individual health selection, for example, by comparing mortality across firms that experienced different levels of

downsizing. These estimates pooled both displaced and nondisplaced workers in a given firm and compared them with workers from firms with different levels of change in employment, suggesting that estimates likely identify a causal effect of job loss on mortality.

The short- and long-term effects reported by Sullivan and Von Wachter are important because they parallel the pattern of short- and long-term effects of job loss on earnings and workers' employment prospects (81, 82). In the short run, displacement is associated with a sharp drop in mean earnings, increased unemployment, and high earnings instability (81). Authors interpret these results as consistent with job loss causing acute stress, which may substantially raise the mortality hazard in the short term. At the same time, in the long run, job loss leaves a scar on workers' careers and earnings, which remains for years even if they return to the labor market (39–41, 82). This is consistent with their evidence of long-run effects of job loss on mortality even decades after job loss.

Studies of this sort based on registry data have also been conducted in the Nordic European countries. A recent study (25, 83) examined the impact of job loss due to establishment closures on nonfatal health events. This was achieved by linking employee-employer registry data to identify job loss due to all establishment closures in Sweden in the period 1987–1988 to subsequent 12-year hospital discharge diagnoses. This study shows that job loss significantly increases the risk of hospitalization due to alcohol-related conditions in men and women, and traffic accidents and self-harm among men only. There was no evidence, however, that job loss increased risk of cardiovascular diseases such as myocardial infarction and stroke, but estimates were imprecise for these outcomes (25, 83). Following a similar approach, Browning and Heinesen (84) assessed whether job loss due to plant closure causes an increased risk of mortality and hospitalization for workers with strong labor market attachment at time of displacement. They used administrative data from Denmark for the period 1980–2006 and applied propensity score weighting and nonparametric duration analysis. Their results suggest that job loss increases overall mortality and mortality from circulatory disease and suicide, as well as death and hospitalization due to traffic accidents, alcohol-related disease, and mental illness.

The studies reviewed above suggest that job loss appears to be associated with an important increase in mortality risk that is particularly high in the first years after displacement but, in some contexts, may lead to long-lasting scars with mortality increases that extend over several years or decades. These findings highlight the importance of considering the long etiologic period between unemployment and mortality, particularly for mortality from chronic diseases such as stroke and heart disease, for which effects may only become evident decades after displacement (43, 49).

To some extent, contradictory findings from the studies above may reflect the impact of policies and institutions. For example, the strong effects of mass lay-offs of the 1970s and 1980s on mortality in Pennsylvania observed by Sullivan and Von Wachter (81) contrast with weaker effects of job loss on mortality during the financial crisis of the early 1990s in Finland reported by Martikainen and colleagues (22). While there are important methodological differences between these studies, on the whole these findings may suggest that the strong social protection policies in Finland may have relieved some of the negative health consequences of mass job loss on health. On the other hand, more recent investigations using a similar approach with data for Sweden (25, 83) and Denmark (84) found that even in these countries with strong social protection systems, job loss due to plant closure is associated with increased mortality. These findings highlight the need for a

new area of enquiry to disentangle the specific institutions and policies that determine the extent to which unemployment influences longevity.

ECONOMIC CYCLES AND HEALTH

Recent decades have witnessed an upsurge of research on the relationship between fluctuations in the economy and mortality, but the history of research in this area expands to as early as the 1920s (85–87). Interest in this question was invigorated by groundbreaking work in the 1970s by Brenner (88–91), who used national time-series data on economic outcome and mortality for England and Wales. Brenner argued that secular declines in mortality over the twentieth century were accounted for by the long-term trend in economic growth over the same period, while fluctuations in mortality rates around the long-term trend ("jumps" up and down around the mortality trend) were explained by increased mortality during recessions and periods of rapid economic growth, and conversely, mortality declines during expansions (89, 90). Despite the large interest in his findings by the public and policymakers, Brenner's work came under severe criticism by many who were unable to reproduce his results and stated that his models were incorrectly specified; that his time-series analyses were susceptible to omitted variable bias (confounding by other time-changing variables); and that his findings were not robust to other time periods and a potential artifact of the choice of years (92–94), among other criticisms.

The field took a complete turn after the publication of influential work by Christopher Ruhm in the early 2000s (37, 95–99). Contrary to earlier reports by Brenner, Ruhm found that economic downturns were associated with reductions in mortality, while economic expansions were associated with rising mortality (37, 95–99). The work by Ruhm did not suffer from the same biases of earlier studies because he exploited variations in unemployment rate fluctuations across US states for the 1972–1991 period in a state fixed effect model, which controlled for state-level differences and other factors at the national level that commonly affected all states. In his pioneer paper, Ruhm (95) found that within-state increases in unemployment rates over this period were associated with significant reductions in total mortality, as well as mortality from 8 of 10 specific causes of death, with particularly large declines from traffic accidents. The only exception was suicide, which was found to increase during economic downturns. Several studies have replicated these findings for mortality from motor vehicle accidents, homicides, cardiovascular disease, and influenza and pneumonia (86, 95, 100).

Over the last 10 to 15 years, numerous studies have been conducted applying the approach pioneered by Ruhm, with several of them finding evidence of *procyclical mortality*—that is, increases in mortality when the economy expands, and mortality declines when the economy contracts (86, 100–108). Not all studies, however, have been able to reproduce these findings, with some studies in European countries, for example, showing no effect of economic cycles or even evidence of *countercyclical mortality*—that is, reductions in mortality when the economy expands. In a recent review, Catalano and colleagues (109) document important gaps in knowledge and contradictory findings across several studies. In this section, we first discuss basic theoretical underpinnings for this hypothesis. We then discuss recent advances in the field and consider several aspects important in understanding the links between the economy

and mortality. In particular, we focus on how the link between economic fluctuations and mortality can differ for rich and poor countries, population subgroups, countries with different social safety net programs, and causes of death. In addition, we discuss potentially contradictory effects of recessions in the short versus the long run as a key element to understand the impact of recessions on health.

RECESSIONS AND HEALTH: SOME THEORY

Epidemiologists will not run short of possible reasons why economic recessions would be damaging to health: Recessions increase the risk of job loss, reduced earnings, marital disruptions, and other undesirable social outcomes, all of which have been extensively shown to be associated with poorer health. The stress mechanism is a common explanation. Declining economic output is likely to lead to stress, which in turn increases the likelihood of experiencing other stressors such as financial difficulties, marital difficulties, parental relationship problems, and poorer well-being (109–112). Anticipation of job loss or financial obligations may also lead to stress and mental health deterioration, even among those who continue to work (109, 113, 114). All of these mechanisms may increase the risk of poor physical and mental health, affecting workers continually employed as well as workers who experience job loss or who are out of the labor market.

On the other hand, economic theory predicts that during times of intense economic activity, individuals may have less flexibility in making time allocation decisions. Insights from the Grossman model discussed in the beginning of this chapter are useful to understanding this mechanism. When the economy improves, leisure time declines; this makes it more costly to make health investments such as spending time exercising and cooking healthy foods. During upturns, the "time price" of medical care increases, as individuals working more hours find it harder to schedule medical appointments. These changes in time allocation flexibility and the time price of medical care would predict that lifestyles will become less healthy when the economy temporarily grows. In addition, as can be predicted from the Grossman model, hazardous working conditions and job-related stress may have direct negative effects on health, particularly for jobs that involve heavy labor. These effects will be stronger during economic upturns given the higher demand for labor and increase in hours worked. Finally, increased economic activity may also lead to greater increases in pollution, transportation, or other unintended consequences of increased economic activity, all of which may pose health risks for both working and nonworking populations.

A SLIPPERY ROAD: CONTRADICTORY EVIDENCE ACROSS TIME, PLACE, AND LEVEL OF ECONOMIC DEVELOPMENT

While most studies based on US data have found that mortality declines during economic downturns, these results have not been reproduced in some European countries, suggesting that welfare state policies or other contextual factors might be important in moderating the effect of recessions.

Using individual-level data and several indicators of economic performance for Sweden, Gerdtham and colleagues (115) found that overall mortality increases during economic recessions among men, while there is no effect among women. This countercyclical pattern for men was consistent for several causes of death including cardiovascular disease, cancer, and suicide, while there was no effect for other causes of death. Svensson (116) extended these findings using regional fluctuations and found that economic recessions are associated with increased incidence and mortality from acute myocardial infarction among prime-working-age Swedish men (ages 20 to 49).

Recent evidence suggests that the well-documented relationship between business cycles and mortality for the United States may not hold for more recent periods. In a recent paper, Ruhm (117) used data for the period 1979–2009 and found that the relationship between the macroeconomic conditions and some causes of death reported in his earlier work (37, 95–99) has shifted and essentially become null using more recent data (117). Ruhm's findings suggest that the relationship between economic cycles and mortality is unstable over time and unlikely to be robust for short periods of 15 to 20 years. While he found that cardiovascular disease and transport accidents continue to decrease when the economy contracts, mortality from cancer and some external causes such as accidental poisoning seems to increase during economic downturns in recent years. On the one hand, these findings suggest that the impact of macroeconomic conditions might be strongly dependent on the context. For example, use of medications may have become more common in response to mental health problems in recent years, leading to a countercyclical relationship between business cycles and accidental poisoning (117). On the other hand, these findings may reflect the instability of fixed effect modeling approaches and their sensitivity to misspecification, for example, if appropriate lag periods between economic conditions and mortality outcomes are misspecified (69).

Evidence from other studies also suggests that the association between business cycles and mortality is far from stable across times and space. Using data from 1800 to 1998 for Sweden, Tapia Granados (104) shows that year-to-year economic growth was strongly associated with mortality declines in Sweden in the first half of the nineteenth century but that this association became increasing weaker in the next hundred years. By the second half of the twentieth century, there was a negative lagged association, with economic growth predicting mortality increase with a short lag of one or two years. A potential explanation is a shift in the composition of causes of death; while economic downturns may have increased mortality from poverty-related conditions such as infectious diseases in the earlier periods, in more recent years, increased industrial activity may increase incidence of conditions associated with affluence such as cardiovascular disease, traffic accidents, diabetes, and cancer. Based on data for England and Wales (118), Tapia Granados also found that while there is a negative short-term relationship between economic growth and life expectancy at birth over the period 1840–2000, this association was much stronger from 1900 to 1950 than from 1950 to 2000, and was very weak in the nineteenth century.

The idea of a shift, or context-dependence, in the relationship between business cycles and mortality is also supported by evidence from studies in low- and middle-income countries, which often find either no or countercyclical associations between the economy and mortality. Using data for Mexico, Gonzalez and Quast (119) found evidence that while in the most developed regions mortality due to noncommunicable diseases increases during economic expansions, in the least developed regions, noncommunicable and infectious disease mortality decreases as the economy expands. They concluded that the association between economic cycles and mortality

may vary by level of development (119). While overall mortality appears to be procyclical in Mexico for those aged 20 to 49, the effects of business cycles differ by cause of death in a way that contradicts findings for high-income countries, with deaths from some causes such as cancer declining as the economy grows, and from other causes such as suicide and homicide increasing when the economy expands (120). Evidence from the United States suggests that infant mortality declines during periods of high unemployment (108), but results for India contradict these findings and suggest that rural infant mortality increases during economic recessions (121). Authors attribute this finding to the fact that recessions stimulate distress among working mothers in India, while in richer countries recessions discourage labor market participation among mothers (121).

In a recent review of the literature, Suhrcke and Stuckler (122) hypothesized that higher average wealth levels and social safety nets may help cushion the impact of recessions in high-income countries, while in low-income settings, where a large share of the population lives below the poverty line, recessions may push a substantial part of the population below subsistence levels. Further research is needed to test this hypothesis, but this could offer an explanation for the discrepancy of results across rich and poor countries or regions.

BUSINESS CYCLES, MENTAL HEALTH, AND SUICIDE

Evidence supports (100, 103, 105) the hypothesis that mental and physical health respond differently to economic shocks. In particular, several measures of mental health seem to consistently worsen during economic recessions and improve during economic expansions. Most importantly, studies have consistently found that suicide increases during economic downturns and declines when the economy improves (95, 100, 102, 103, 105, 123–127). This association, however, may also vary across different regions and countries, with not all studies reporting countercyclical suicide mortality (86, 106, 128–130). Worsening of other mental health outcomes such as depression has also been reported (130–132).

The difference between studies on physical and mental health reflects important aspects of the mechanisms through which recessions may influence health. Unlike medically diagnosed physical conditions—many of which take years or decades to develop—suicide and mental health can change in the short term in response to sudden shocks. The fact that effects of recessions are most consistent for these outcomes, therefore, is generally in line with the stress mechanisms discussed above.

RECESSIONS AND HEALTH-RELATED BEHAVIOR

Studies exploring potential mechanisms linking recessions to health have focused primarily on individual behaviors such as smoking and alcohol consumption (36, 37, 99). Studies assessing this question typically link national or regional economic indicators to individual-level survey data on behavior. Many of the studies conducted seem to be in line with the hypothesis that people behave in more healthy ways during harsh economic times, while they tend toward less healthy behavior when the economy is doing well. For example, Ruhm used US micro data to show that smoking,

alcohol consumption, and excess body weight decline during recessions, while leisure-time physical activity increases (37, 99).

In an interesting recent study, Xu (133) combined data from the Behavioral Risk Factor Surveillance System (BRFSS) and the National Health Interview Survey (NHIS) with data on employment from the Current Population Survey (CPS) in the United States to assess the effects of changes in wages and working hours caused by business cycles on health behavior among low-educated persons. In line with Ruhm's (95) findings, increases in wages and working hours caused by economic expansions were associated with higher consumption of cigarettes. Increases in working hours during expansions were also associated with less physical activity and physician visits. Their models suggest that most of these changes are driven by shifts in employment status, rather than changes in the wages and hours worked by those who are employed throughout. In general, these findings are in line with the hypothesis that shifts in the opportunity costs of time may be key to understanding short-term health variations associated with the business cycle in the United States.

These findings, however, have not been replicated in some studies outside the United States, as well as some US studies, several of which dispute the idea that behavior improves during recessions. Using micro data for Finland for the period 1978–2002, Bockermen (134) found that an improvement in general economic conditions decreases body mass index, which does not support evidence of the opposite pattern for the United States. Using survey data for Canada, Latif (135) found that a rise in the unemployment rate increases average body mass index and the probability of being severely obese. Examining how business cycles influence eating habits, a recent US study (136) found that a higher unemployment rate is associated with reduced consumption of fruits and vegetables, increased consumption of unhealthy foods such as snacks and fast food, and overall declines in the share of total food representing healthy foods. These findings contradict earlier expectations that individuals would behave better during recessions, and, on the contrary, suggest that a poor economy is associated with a poorer diet.

Equally conflicting evidence has been put forward for alcohol consumption. While per capita alcohol consumption increases during economic expansions and decreases during economic recessions, survey data suggest that the prevalence of heavy drinking increases during economic recessions. Using data from 1984–1995 Behavioral Risk Factor Surveillance Surveys (BRFSS), Dee (137) shows that the prevalence of binge drinking increases substantially during economic downturns, even among those who remain employed. Using BRFSS data for 2003–2010 in the United States, Nandi (138) found that increases in unemployment are associated with fewer drinks consumed in the past month and decreased prevalence of past-month heavy drinking, but no changes in other health behaviors. Based on micro data for Finland, Johansson (139) found that alcohol consumption increases during expansions, while the probability of being a drinker remains unchanged. Lo (140) used data from the 1997–2011 Combined National Health Interview Survey and found that rising unemployment is associated with higher prevalence of heavy-drinking frequency, but lower prevalence of heavy-drinking quantity. These studies would seem to suggest that harmful alcohol consumption increases when the economy worsens, but decreases when the economy improves.

In summary, while much of the evidence presented in the initial studies by Ruhm and others suggested that health behavior improves during recessions, more recent evidence contradicts these findings and suggests that harmful forms of behavior appear to be common during economic

downturns. Nevertheless, contradictory findings may again reflect the differential impact of business cycles on behavior across different countries and institutional settings, a question that remains yet underexplored in the current literature.

WHO SUFFERS DURING RECESSIONS?

Studies on the impact of economic downturns on mortality typically rely on aggregate data, ignoring potentially heterogeneous effects across individuals with different vulnerability to negative labor and social outcomes shocks. Aggregate associations might mask differential effects of economic downturns across workers who remain in the labor market versus workers who lose their jobs. Effects of economic downturns on the health of low-skilled, low-wage workers may also differ from the effect on higher-skilled workers with more stable careers.

There have been few studies addressing this question in the literature. However, recent evidence from the United States suggests that the impact of economic recessions on labor market outcomes is borne disproportionally by men, black and Hispanic workers, youth, and lower educated workers, partly due to the demographic composition of workers across industries and occupations which are differentially affected by economic downturns (141). Few studies have examined whether this results in heterogeneous health effects. A study based on US data found that working-age adults with low educational level, as well as employed workers, suffer increased mortality during economic contractions, while the more highly educated, unemployed, disabled, and retired experience reduced mortality during economic downturns (103). However, this study was based on a one-time cross-sectional assessment of demographic characteristics. Fontenla, Gonzalez, and Quast (142) partly addressed this caveat by using panel data disaggregated by county and found that mortality rates for whites and Latinos decrease when the economy contracts, while there is generally no relationship for blacks. They also found that mortality is more procyclical for counties that are less racially/ethnically diverse. All together, these findings suggest that the impact of business cycles on mortality may vary by race and ethnicity.

Changes in employment occurring alongside economic downturns may lead to fundamentally different effects on mortality. Displaced workers may experience a drop in income and lack of access to health services, which may in turn lead to increased mortality during an economic downturn (109, 143). Employed workers may experience increased stress due to downsizing or other stressors (144, 145) leading to poor mental health, but they may also smoke and drink less due to financial constraints during economic downturns (36, 37). The trade-off between negative and positive spillovers to health may depend on the composition of the population and the strength of effects in multiple directions.

Another important consideration is age. Evidence for the United States suggests that economic downturns are associated with mortality declines among children, working-age adults (ages 25–59) and older people (ages 60+) (105). Among middle-age working-age adults, the majority of deaths contributing to declining mortality during economic upturns appear to be from traffic accidents and not from causes of death presumably related to work stress such as heart disease and stroke (105). While the effects of recessions are stronger in working-age populations, because the majority of deaths in general occur at older ages, most of the deaths contributing to procyclical mortality occur at old age.

Taken together, these results suggest that fluctuations in the economy may impact health partly through mechanisms that are not directly related to labor market participation. For example, recessions may influence family and co-residence arrangements in ways that may end up benefiting older adults. Evidence suggests that, as a result of the recent recession, the prevalence of vertically extended family arrangements has grown (146, 147). Economic downturns may increase the likelihood that older people share a household with their children or other family members, increasing their social contacts, social participation, and well-being and reducing their risk of illness. Others have argued that business cycles might influence the quality of health and nursing care, which are heavily used by older people. When the economy is good, hospitals and nursing homes may experience staffing shortages as skilled workers go to work in other sectors of the economy, potentially diminishing health care inputs for the elderly (105).

LONG-TERM VERSUS SHORT-TERM EFFECTS OF RECESSIONS ON HEALTH

Much of the literature has focused on cyclical short-term fluctuations in mortality in relationship to business cycles. A potential limitation of this approach is the lack of understanding of how exposures across the lifecourse may influence the development of disease over long and complex etiologic periods. For example, while recessions may lead to short-term changes in health behavior, over the long-term, exposure to recessions in key periods of the lifecourse may lead to permanent changes in health and aging over the long run.

An example are studies analyzing the effects of economic conditions around birth on late-life mortality, often showing that being born under adverse economic conditions may have long-lasting negative health consequences (148, 149). In these studies, the period around birth is considered critical, with maternal malnutrition or other negative exposures during pregnancy leading to developmental disadvantages manifesting as increased risk of chronic conditions in later life (150). Whether recessions in critical periods in early or late adulthood influence health outcomes later in life, however, has been less explored.

Evidence suggests that recessions experienced around the transition from school to work, another critical lifecourse period (12, 13), may trigger a pathway toward cumulative disadvantage characterized by less favorable and unstable labor market trajectories (2–4), which may ultimately lead to poorer health later in life. Two recent studies have explored whether leaving school in a bad economy has long-lasting effects on health. Maclean (151) linked data from the National Longitudinal Survey of Youth to state unemployment rates in the month and year of leaving school between 1976 and 1992. Results suggest that men who left school at times of high unemployment have worse health at age 40 than otherwise comparable men, while for women, leaving school in a bad economy was associated with less depressive symptoms at age 40.

Hessel and Avendano (152) used data from the Survey of Health, Ageing and Retirement in Europe (SHARE) to examine whether national economic conditions at the time of leaving school or college were associated with physical functioning at ages 50 to 74. They linked national unemployment rates in the year of leaving school for cohorts completing education between 1946 and 1986 to SHARE data for 13 European countries. Contrary to Maclean's results, they found men leaving school during times of high unemployment had lower levels of physical impairment than otherwise

comparable men, but women graduating in poor economic times had worse health and physical function in old age. Exploring potential explanations, they found that economic conditions at the age of leaving school were associated with different labor market, marriage, fertility, and health behavior outcomes, which may provide a partial explanation for the negative effects on women's health.

Following a similar approach, Leist and colleagues used data from SHARE for 11 European countries to assess whether recessions experienced between ages 25 and 49 were associated with cognitive function at ages 50 to 74. Their results suggest that recessions experienced at ages 45 to 49 years among men and in the early-to-mid career stages among women (25 to 44 years) are associated with poorer later-life cognitive function, possibly through more unfavorable labor market trajectories. Although preliminary, these findings suggest that policies that ameliorate the impact of recessions on labor market trajectories may promote a healthier later-life cognitive function.

Another critical period is the transition to retirement. Recessions in the years leading up to retirement can have a critical impact on the timing and circumstances in which older workers decide to leave the labor market. The relatively short-time horizon in the labor market of a laid-off worker approaching retirement may reduce their chances of re-entering the labor market (153). As a result, older workers may be forced to accept lower wages, permanently leave the labor market, or collect social security benefits early (153, 154). This leads to major losses in both present and future income (155–157), which may have devastating consequences for older workers' retirement plans; increase their risk of poverty; and ultimately diminish their financial, physical, and psychological well-being (153, 158–162).

In a recent paper, Coile and Levine (163) examined whether these changes translate into long-run effects of recessions around retirement on mortality. They used vital statistics mortality data between 1969 and 2008 to obtain age-specific cohort survival probabilities at older ages, and linked these to labor market conditions at earlier ages. Their results indicate that experiencing a recession between 50 and 61 leads to a substantial reduction in longevity. Exploring the potential explanations, they used survey data to show that recessions at ages 50 to 61 lead to several years of reduced employment, health insurance coverage, and health care utilization, which may partly contribute to lower survival. An important finding is that these effects may be less important beyond age 62, the age at which older Americans are entitled to collect social security benefits.

In summary, these studies suggest that a narrow approach that focuses on the short-term effects of business cycles on health ignores the potential long-term health consequences of economic downturns experienced at key stages of the lifecourse, including the transition from school to work and from employment to retirement. These findings highlight the importance of adopting a lifecourse perspective to understand how recessions across the lifecourse influence the development of health and disease in the long run.

UNDEREMPLOYMENT, JOB INSECURITY, AND HEALTH

Changes in employment law since the 1970s have brought changes in job security and contractual arrangements for many workers (164, 165). The proportion of workers in fixed-term contracts

has increased during the last three decades, although these patterns have occurred at different pace in different countries. For example, while in Spain one-third of all employment contracts are currently fixed-term, in Germany, fixed-term employment remains relatively rare and represents about 8% of all contracts (166). As follow-up to these developments, evidence has emerged during recent years documenting an association between job insecurity—sometimes referred to as "precarious employment"—and poor physical and mental health, with some findings suggesting that negative health effects of job insecurity might be comparable with those of unemployment (164, 167–173).

Studies on job insecurity can be classified into two categories: A first group of studies examines the impact of perceived job insecurity on health. A second group of studies examines how exogenous shocks to job security, generally due to downsizing and workplace closure, impact health. We separately discuss each of these two types of studies.

PERCEIVED JOB INSECURITY

Typically, studies of this type assess the statistical association between self-report of job insecurity and a measure of health. A study based on the British Whitehall Study II is a good example of studies in the first category. Ferrie and colleagues (174) asked participants how secure they felt in their present job in 1995/96 and again in 1997/99, and compared the health of participants who reported a change in job security or were insecure in both periods with the health of participants who reported job security in both periods. They found that loss of job security was associated with worse self-rated health and minor psychiatric morbidity. Those who lost job security had higher morbidity than those who were secure in both periods. However, those who gained job security also showed worse negative psychological health than those who were secure in both periods. Those exposed to chronic job insecurity—defined as job insecurity in both periods—had the worst health profile. Effects were also observed in some but not all measures of physical health, including blood pressure and body mass index. Similar studies have found comparable results for a wide set of measures of physical and mental health (168, 175–183).

Disentangling confounding from mediating factors remains challenging in studies assessing the impact of job insecurity. In a recent systematic review and meta-analysis, Virtanen and colleagues (184) combined estimates from individual-level data from 13 cohorts with studies identified through a systematic review to obtain pooled estimates of the impact of job insecurity on coronary heart disease. After controlling for confounders, a significant but modest association between job insecurity and coronary heart disease was observed. They concluded that this modest association is partly attributable to poor socioeconomic circumstances and less favorable risk factor profiles among people with job insecurity. The fact that job insecurity remains only marginally associated with coronary heart disease after full control adjustment brings up the question of residual confounding by socioeconomic status. On the other hand, controlling for these factors may also lead to overadjustment if they are mediators of the association between job loss and coronary heart disease.

Despite the potential of these studies and the strength of their longitudinal design, most of these studies are vulnerable to bias due to selection or confounding, even if assessments are longitudinal. Changes in job security may be accompanied by changes in other relevant factors

associated with health. While it is plausible that job insecurity triggers negative health effects (185, 186), it may also be the case that less healthy workers more often end up in less secure jobs (166, 187). Similarly, workers with chronic job security are likely to be a nonrandom sample of the population different from the set of workers with secure jobs. Understanding the causal nature of this association has important policy implications: If selection is the dominant mechanism, policies to help less healthy workers find stable employment should be the focus of interventions. In contrast, if job insecurity does have a causal effect on health, policies to improve contractual arrangements, or social protection programs to insure vulnerable workers against the risks of job insecurity should be the focus of interventions.

DOWNSIZING STUDIES

To overcome the limitations of studies on perceived job insecurity, a second set of studies examines whether the health of individuals in workplaces that experienced downsizing differs from the health of workers in otherwise comparable workplaces that did not experience downsizing. The rationale is that downsizing generates exogenous job insecurity for workers who continue to be employed in a given workplace. In addition, these studies directly tell us what the impact is of potential downsizing polices as a firm or government policy. In a classical study, Vahtera and colleagues (188) used data from Finland to assess whether organizational downsizing had an effect on sick leave. They found a significant association between downsizing and sick leave, with the risk of absenteeism being more than twice as large after major downsizing as compared with minor downsizing. Similar effects have been observed for the use of psychotropic drugs (144). Potential mechanisms associated with downsizing include negative changes in job control, impaired support from spouse, and increased prevalence of smoking (145).

The evidence presented above, however, is to a large extent based on studies of Finnish municipal employees; the extent to which these results can be generalized to other contexts is uncertain. Other studies, for example, have found less evidence of effects of downsizing on health for "surviving" employees. Osthus (189) used conditional fixed effect models to assess whether downsizing has an effect on sickness absence based on national registry data for Norway, and found evidence that downsizing leads to slightly less sickness absence among workers who remain employed, but seems to have little negative short-term effects on health. Following a similar approach, Osthus and Mastekaasa (190) used survey data for 1997–2003 and found no evidence that downsizing has any impact on the health of workers that remained employed. Two Swedish studies reported health deterioration for workers in companies experiencing downsizing (191) and organizational instability (192). However, other studies found mixed or no effects of downsizing on health in Sweden (193, 194).

Another set of studies examines the health impact of government privatization policies on workers. In a review of the literature, Egan and colleagues (195) identified eleven studies dating from 1945 on the impact of privatizing industries and the health of employees in the public sector. They included only experimental and quasi-experimental studies from countries in the Organization for Economic Cooperation and Development (OECD) member countries. They found that the most robust effects were on measures of stress-related illness among employees

after privatization, including company downsizing. However, they found no robust evidence of an effect of privatization on the risk of injuries, and conclude that there is insufficient evidence on the health effects of privatization.

In conclusion, there are many theoretical reasons why job insecurity might have a negative impact on the health of workers. Studies on perceived job insecurity do show an association with poorer health, but there is much contradiction between quasi-experimental studies examining the impact of downsizing and privatization, with several of them finding no clear health effects of downsizing. More research is needed to identify the specific conditions under which downsizing may be detrimental for health. For example, effects may differ across public and private sector companies. In addition, most evidence comes from the Scandinavian countries, characterized by low unemployment, strong unions, and extensive worker protection legislation (189). Further research is needed to assess whether downsizing has stronger and more consistent effects in countries without the institutional features characterizing Nordic European countries (194).

UNDEREMPLOYMENT AND HEALTH

Most studies on the impact of employment status essentially distinguish a broad category of employment from several forms of nonemployment. It has been argued (32), however, that such broad conceptualization of employment insufficiently considers the variety of potential forms of "inadequate employment." The health effects of several forms of inadequate employment have been primarily examined by Dooley and Prause (32), who refer to *underemployment* as the concept encompassing these new forms of inadequate employment, based on the Labor Utilization Framework (196, 197).

Over the last decades, particularly in the United States, low unemployment rates have been paralleled by a rise in the proportion of workers in "disguised unemployment." This term, first coined by Robinson (198), has no official definition but typically refers to several forms of "inadequate" employment based on hours worked (involuntary part-time work) and wages (poverty-level pay). It includes workers who involuntarily work fewer hours, workers under low pay, and discouraged workers who involuntarily turn permanently out of the labor force given the poor prospects of finding employment. Another important aspect of underemployment is the mismatch between levels of education and occupation (32, 197). According to this perspective, underemployment carries major health risks similar to those of unemployment. Based on this perspective, Dooley and Prause (32) argued for a new definition of employment as a continuum, with underemployment falling somewhere between adequate employment and unemployment.

Several studies have been conducted to examine the longitudinal association between underemployment—or a specific component of underemployment—and health. In a special issue on this theme, Friedland and Price (199) used data from a US national representative sample of adults of working age to assess the effects of underemployment using a longitudinal design that controls for prior health. Their results suggest that underemployed workers report lower levels of health and well-being than workers with "adequate" employment, but relationships vary by type of underemployment and indicators of health and are sometimes modest. They found an association

with underemployment for four of the seven indicators used, with low wage generally being associated with poorer health. However, they found less clear evidence of an association between "hours-underemployment" and physical health; it predicted lower psychological well-being but higher levels of job satisfaction.

An important part of the literature on underemployment has focused on relatively young workers, who may be more vulnerable to underemployment. Dooly and Colleagues (200) used panel data from the National Longitudinal Survey of Youth to examine the impact of transitions from adequate to inadequate employment on depression from 1992 to 1994. Controlling for baseline depression and a wide range of confounders, they found that transitions to underemployment were associated with significant increases in depression, which remained after controlling for income, marital status, and job satisfaction. In examining a possible selection mechanism, they found that prior depression did not predict risk of becoming underemployed although it did predict risk of unemployment, particularly for lower-educated workers.

Dooly and Prause (201) found evidence on their work that transitions from adequate employment to underemployment are associated with lower self-esteem at young age; increased symptoms of alcohol abuse in early adulthood; and onset of depressive symptoms in workers' late twenties and early thirties. In one on their papers (201) they explicitly examined the role of health selection more in general. They concluded that several indicators of mental health, including depression, were associated with subsequent underemployment, and they highlighted the importance of controlling for selection in these studies.

Findings from the studies above suggest that not only transitions to unemployment but also transitions to underemployment are associated with worse psychological health. While sources of reverse causation and confounding remain a concern, these studies offer a promising avenue for future research. A potential concern with the definition of underemployment is that it encompasses many elements, making it difficult to distinguish which specific aspects of underemployment may be detrimental to health, and which ones may be amenable to policy. For example, whether involuntary reductions in the number of hours worked are detrimental to health independently of changes in wages or occupational mismatch would offer unique opportunities for policy interventions that cannot be easily envisaged in studies compiling a measure of underemployment combining multiple measures. In addition, while there is some evidence of mental health effects, there is no convincing evidence as yet that underemployment is consistently associated with physical health.

An important consideration for future studies on how underemployment influences health is the potential to use recent changes in employment laws as a natural experiment to disentangle causation from selection mechanisms. For example, few studies have examined how changes in the length of the workweek legislation in some European countries such as the "Aubry" Law in France, affected the well-being and mental and physical health of workers. In addition, few studies have examined how changes in the regulation of contractual arrangements and minimum wage laws, which may have had an impact on underemployment, affected the health of workers affected by these reforms. Not only do quasi-experimental studies help us to disentangle to what extent the association between underemployment and health is causal, but they will also point to the potential health impact of specific labor market policies and laws.

EMPLOYMENT PROTECTION POLICIES AND HEALTH

In this section, we focus on two specific areas of employment policy that are aimed at increasing job security and ensuring long-term employment. Policies on maternity leave and retirement have attracted attention recently, especially in European countries, where social protection policies during times of demographic change are conceived of as central to successful societies and labor markets. How each of these policies may influence health is complex, since it is likely that such policies influence a broad set of outcomes and work through multiple channels. We discuss current research on these two areas to illustrate the challenges as well as the potential in studying the health effects of these policies.

MATERNITY LEAVE POLICIES

Much has been written about the impact of unemployment on health, but less is known about how other forms of leave from employment relate to health. Parental leave, and particularly maternity leave, may be especially relevant to the health of working mothers. During the second half of the twentieth century, high-income countries witnessed a remarkable increase in female labor force participation. Women with children were no exception to this pattern. In the United States, for example, 64% of mothers with children under the age of 6 were in the labor force in 2011, as compared to 33% in 1975 (4, 202). In response to these trends, some countries enacted comprehensive maternity leave policies during the second half of the twentieth century to help households with children cope with the competing demands from work and family.

Maternity leave legislation extends women the right to take a period of leave from work around childbirth and, in many countries, to receive income support compensation during maternity leave. Initially, maternity leave policies were motivated by concerns about the health of the child and the mother in the period around birth. Since the end of the 1960s, however, maternity leave has also become a job-protected period out of work to care for newborns and young children. Recent evidence suggests that by protecting employment among mothers in the period around birth, maternity leave leads to better long-term labor market outcomes after maternity including wage level and growth, career prospects, labor market attachment, and employability (1–5). An important question, therefore, is whether these policies also affect the health of mothers and their children both in the short and the long run.

There are several reasons why maternity leave policies may lead to better health among mothers and their children. Ruhm (14) argues that the health of young children depends on the "health stock," the level of medical technology, the price of and access to health care, household income, and time investments of parents. Among all these mechanisms, he argues that time investments of parents might be of particular importance to the health of children. Raising children is a time-intensive activity, and time investments even before birth—in the form of better nutrition and prenatal care—are likely to deliver better outcomes for children in the short and long run. An example comes from the literature showing high returns of investments in breastfeeding behavior on children's cognitive development (203). Similarly, the time available during the early weeks

of life may be crucial to the prevention of accidents or other health problems with long-lasting consequences.

Evidence of the impact of maternity leave legislation on the health of children comes from two sources. First, cross-national comparative studies have examined whether differences in reforms introduced over the last decades affected the health of children across high-income countries. In a pioneer study, Ruhm (14) used aggregate data to assess the impact of parental leave laws on infant mortality in 16 European countries over the period from 1969 to 1994. He used a country fixed effect model that related the weeks of parental leave entitlement to infant mortality over this period. While based on aggregate data, an advantage of this approach is that exposure is based on "exogenous" changes in legislation on the number of weeks that affect all women and births that happened to occur after the introduction of the law. By comparing affected and unaffected cohorts within each country, the study is able to examine the overall effect of legislation on infant mortality. His results provide one of the most important pieces of evidence suggesting that more generous paid parental leave substantially reduces deaths of infants and young children. Evidence was stronger for postneonatal mortality and child fatalities than for perinatal mortality, neonatal deaths, or low birthweight. Rights to a year of job-protected paid leave were associated with around 20% declines in postneonatal deaths and 15% decline in fatalities between the first and fifth birthday.

Based on a similar design, Tanaka (16) used aggregate data to examine the effects of job-protected paid and other leave on child health outcomes from 1969 to 2000 in 18 OECD countries, including Japan and the United States. Consistent with the earlier work of Ruhm, he found that extension of weeks of job-protected paid leave has significant effects by decreasing infant mortality and also by improving birthweight outcomes. Importantly, he did not find this effect for unpaid maternity leave, suggesting that lack of adequate payment and job protection during leave will not deliver the same health benefits.

A review of the literature by Staehelin and colleagues (15) examined 13 original studies examining whether maternity leave policies influenced the health of children and their mothers. They found that a positive association has been found between the length of maternity leave and the duration of breastfeeding. Maternity leave was also associated with lower perinatal, neonatal, and postneonatal mortality and child mortality. They pointed out, however, that these findings come primarily from "ecological" studies, while there is less evidence of impact on other health outcomes. In a separate review, Ruhm (4) discussed some of the potential mechanisms through which maternity leave may improve health outcomes. Among these, maternity leave is a potentially promising explanation, given the potential benefits of breastfeeding on child health. To assess this question, a study by Baker and Milligan (204) examined the effect of a major increase in maternity leave mandates in Canada by comparing the health of mothers giving birth before 31 December 2000 and entitled to a maximum of approximately six months of job-protected compensated maternity leave, to that of mothers to children born after that date, whose benefit entitlement and job protection benefits were extended to about one year (205). Results suggest that extended maternity leave mandates led to large increases in the attainment of critical breastfeeding thresholds, although they find no effect of mandates on most indicators of child and maternal health. Recent evidence from a different policy provides some indication of the potential impact of providing time for parents to breastfeed. Using differences in the enactment of breastfeeding laws across US states, Hawkins (206) found that breastfeeding initiation was 1.7% higher in states

with new laws to provide break time and private space for breastfeeding employees, particularly among Hispanic and black women.

In conclusion, the studies above provide convincing evidence that paid maternity leave benefits have the potential to improve child health and reduce infant mortality. Further research is needed, however, to understand why only paid benefits seem to impact health, and to determine the specific number of weeks of maternity leave that would optimize child health.

MATERNITY LEAVE AND THE HEALTH OF MOTHERS

As discussed in the initial sections of this chapter, maternity leave may also improve the health of mothers by diminishing the human capital loss associated with a period out of the labor market around childbirth. Maternity leave increases job protection and female labor market attachment by enabling women to return to their employer after a short period of leave (1, 4, 12), increasing job continuity, and preventing the erosion of firm-specific skills. By protecting mothers' career prospects, earnings accumulation, and labor market attachment (1–5), maternity leave may improve the socioeconomic circumstances of mothers, with potential long-lasting health consequences.

Research on the health impact of maternity leave legislation has primarily focused on health in the years around birth. In a systematic review, four out of six studies reported a positive association between the length of maternity leave and mental health in the postpartum period (15). A study using cross-sectional variations in maternity leave entitlements across US states before 1993 found that mothers entitled to 8 to 12 weeks' leave after childbirth had fewer depressive symptoms than women entitled to only 6 weeks (207). Another study found that maternity leave of 12 weeks or longer is associated with lower depression scores in women with marital concerns, and lower depression and less anger among women with low work rewards, compared to maternity leave of 6 weeks or less (208). Two studies found that general mental health at 7 and 9–12 months after childbirth, measured by depression, anxiety, general positive affect, and life satisfaction, was better in women with maternity leaves beyond 15 weeks and 24 weeks, as compared to women with less than 9 weeks of leave (209, 210). No effects were found in the other two studies included in this systematic review.

The study by Baker and Milligan (204) referenced above on the maternity leave mandates in Canada, which extended benefit entitlement and job protection benefits to about one year, found no evidence that these extensions led to improvements in maternal health. In a separate study, Chatterji and Markowitz (207) used variations in state maternity leave entitlements before the introduction of the federal level by the Family and Medical Leave Act (FMLA) in 1993. They found that maternity leave shorter than 12 weeks, as well as paid maternity leave shorter than 8 weeks, were both associated with increased depressive symptoms. In addition, paid maternity leave shorter than 8 weeks was associated with poorer overall self-rated health.

There are three important considerations in regard to these studies. First, the strength of these studies is the focus on the impact of maternity leave policies, rather than associations between the length of leave and maternal health. This focus enables us to understand the potential benefits of introducing a specific policy, but it also improves causal inference by comparing cohorts that "happened to" experience different policy regimes. While the evidence is not all consistent, there would

seem to be some indication that maternity leave policies can improve mental health in the period after birth, with less clear benefits on physical health outcomes. A second consideration is the fact that most studies on this area have been based on data for North America. Important to note is the fact that in the United States, maternity leave benefits are unpaid, while in many European countries generous paid maternity leave benefits are offered to working mothers. Whether paid maternity leave benefits will show stronger effects remains to be explored in future studies.

A final key consideration is the fact that most research has focused on the impact of maternity leave on the health of mothers in the period around childbirth. However, the mechanisms through which maternity leave might influence a mother's health are likely to operate in the long run. The impact of protected maternity leave on a mother's wage level and growth, career prospects, labor market attachment, and employability (1–5) is likely to bring health benefits only measurable as women reach older ages and face the cumulative toll of exposures over the lifecourse. The current literature, however, narrowly focuses on the short-term benefits of maternity leave. The crucial question is whether a policy that potentially improves the socioeconomic position of women and reduces gender differences in labor market trajectories also generates health capital. This line of enquiry offers a promising avenue for research on the long-run effects of parental leave policies on working mothers' health and, increasingly, on the health of working fathers as well.

RETIREMENT POLICIES AND HEALTH

Increases in life expectancy during the last decades, coupled with declines in the age at which workers retire in many high-income countries, has sparked interest on the impact of retirement policies on health. With the prospect of many countries considering or implementing policy reforms to increase retirement age, the question of whether retirement influences health is high on the policy agenda. Retirement is an important labor market transition that brings large changes to an individual's life. While there is little doubt that poorer health is associated with early retirement, whether the transition to retirement is associated with changes in health remains a matter of debate. Studies examining this question so far have reached no consensus on whether retirement promotes or harms health. As pointed out earlier (211), however, many of these studies have faced a number of problems, as they are unable to distinguish the effects of aging from those of retirement. Many of the descriptive studies lack an adequate control or comparison group; this is important because the decision to retire is not random but may be associated with important changes potentially correlated with health, and leading to retirement. In addition, the effects of retirement on health may depend on many contextual factors including the adequacy of retirement benefits as well as individual factors such as occupation, socioeconomic status, and marital status.

Over the last decade, we have seen an increased interest in understanding whether retirement has a causal effect on health. This debate has primarily focused on understanding how reforms on the age of retirement—namely the age of compulsory retirement or the minimum age of retirement—might impact health. These laws have a strong effect on retirement decisions: While a substantial proportion of workers retire before statutory retirement age, a higher statutory age of retirement generally encourages individuals to work longer. Rather than reviewing the extensive evidence on the association between retirement and health, in this section we aim to focus on the smaller but increasing set of studies that have attempted to establish causality. Two pieces of

evidence are important in this respect: First, we discuss some of the studies using longitudinal and panel data to assess how transitions to retirement relate to health transitions controlling for potential confounding. Second, we discuss the evidence on the impact of policy reforms on the age of retirement on health. Not only do the latter studies provide a more direct entry point for recent policy reform discussions but they also help us in disentangling the extent to which the relationship between retirement and health may be causal.

RETIREMENT AND HEALTH: SOME THEORY

There are several reasons why retirement may bring benefits to health. As discussed in the first sections of this chapter, workers will invest more in skills and human capital over their working life if they expect to retire later, as the time-horizon for claiming the returns of human capital investments expands. Retirement policies thus influence human capital investment decisions over the working life, such as acquiring on-the-job training, which in turn influence career and earning trajectories over the lifecourse. Based on the Grossman model (6) discussed at the beginning of this chapter, a key issue is the fact that retirement provides individuals with more flexibility for time allocation decisions. After retirement, the opportunity cost of time declines, increasing the amount of time individuals have to invest in their health. For example, after retirement, individuals may have more time to exercise or cook healthier foods. During retirement, the "time price" of medical care declines, making it easier for individuals, for example, to schedule medical appointments. These changes in time allocation flexibility and the time price of medical care would predict that lifestyles will become healthier after retirement. In addition, for many low-skilled workers, retirement may release them from hazardous working conditions and job-related stress.

Retirement, however, may also be detrimental to the health of workers. Retirement may lead to a loss of the nonfinancial benefits of work, such as a time structure for the day, opportunities for social contacts, and self-esteem and status (28), some of which may lead to poorer health. Retirement may also lead to a loss of traction or motivation (18) or generate changes in health behavior. For example, individuals may reduce their level of overall physical activity. Changes in time allocation geared by retirement may also increase opportunities for other unhealthy behaviors, such as smoking and alcohol consumption. These effects would predict that policies that increase retirement age will be beneficial for the health of workers.

DOES RETIREMENT INFLUENCE HEALTH?

Whether retirement has an impact on health remains heavily debated, but some recent studies have started to shed light on this question. Two types of studies have brought new insights over the last decade: longitudinal panel studies and studies examining the impact of retirement age policy reforms.

Longitudinal studies typically follow the health of workers over the years prior to, during, and after retirement and compare it with the health trajectories of workers who do not retire. In a recent study, Westerlund and colleagues (212) examined yearly trajectories of self-rated health

among employees on the French national Gas and Electricity Company (GAZEL) for 7 years before and 7 years after retirement. Although poor health increased with age, between the year before and after retirement, the prevalence of poor self-rated health fell from 19% to 14%, which corresponded to a gain in health of 8 to 10 years, a benefit maintained over the seven years after retirement. This effect was particularly strong for workers reporting a poor work environment before retirement.

Based on a similar design, Mein and colleagues (211) used data from the Whitehall II Study to compare the mental health trajectories—measured with the SF-36 form—of British civil servants who retired, to the health trajectories of continually employed workers. They found that mental health improved after retirement, while it deteriorated among those who continued to work. Contrary to the GAZEL Study, however, mental health improvements were confined to higher-grade workers. Physical functioning declined in a similar fashion among working and retired civil servants. The authors concluded that retirement had no effects on physical functioning, but was associated with better mental health among high-grade employees.

Jokela and colleagues (213) re-examined data from the Whitehall II Study to assess whether effects were dependent on the type of retirement. Compared with continued employment, compulsory retirement at age 60 and early voluntary retirement were associated with improvements in mental health and physical functioning. In contrast, retirement due to ill health was associated with poorer mental health and physical functioning. These findings highlight the important role of health-related selection as a potential explanation of the negative association between retirement and health. Poor health was a predictor of all forms of retirement; however, this would not explain why compulsory or voluntary retirement was associated with better health, and may even suggest an underestimation of the benefits of retirement.

Roberts and colleagues (214) used data from the Whitehall II Study to examine the impact of retirement on cognitive function. This outcome is of particular interest in view of the "use it or lose it" hypothesis, which suggests that age-related cognitive decline can be lessened by engaging in cognitively demanding activities that exercise cognitive functions (215). This would suggest that, to the extent that work is cognitively demanding, retiring later could bring benefits to cognitive function. Studying cognitive function longitudinally is challenging, however, because of practice effects: Individuals learn with practice and improve their cognitive test scores with repeated assessments. In their study, all workers experienced an improvement with cognitive function over time. However, those who retired experienced smaller improvements in mean cognitive test scores than those who continued to work, although these changes were not significant for most cognitive test scores. Their results suggest that cognitive function may show little change or, if anything, a slight deterioration after retirement.

The studies above offer important insights because they follow participants over time and compare within-individual changes before and after retirement, thus controlling for potential confounding due to differences across individuals. Many of the studies, confined to European populations, cast doubt on the notion that retirement is bad overall for health: The prevailing finding appears to be that in the short-term, retiring is associated with an improvement in mental health; retirement might have little or no benefits on physical health, but there is no clear evidence that it harms physical health. Because the selection hypothesis would generally bias results toward a detrimental effect of retirement on health, due to the higher risk of retirement due to illness, selection seems an unlikely explanation for these findings.

There are, however, two limitations to the studies discussed above. First, they both deal with a selection of workers in Britain and France who have worked under relatively favorable physical and contractual conditions. In addition, in the case of Whitehall British civil servants, the sample is composed of white-collar workers. Retirement might have different effects, for example, for low-skilled workers in less favorable physical and contractual conditions. Second, estimation in these studies is based on a comparison of individuals who retire at a given point versus those who continue to work; these comparisons may still be confounded by, for example, new events that motivate individuals to retire early (e.g., the death of a spouse, a change in family relations).

To overcome some of these limitations, a group of new studies have used quasi-experimental methods to assess the impact of retirement on health. Most notably, these studies use differences across cohorts in eligibility to retirement benefits based on retirement legislation on statutory retirement or pensionable ages. Charles (216) used policy variation in mandatory retirement and Social Security benefits that influence retirement incentives by age and cohort in the United States to examine the impact of retirement on depression. While he found a negative correlation between retirement and depression, after using policy variations across cohort and age discontinuities as instrumental variables, he found that retirement in fact leads to better mental health and well-being. A series of studies have used a similar approach that exploits country-level variations in the age of eligibility for early and full retirement benefits in the public pension system as instrumental variables. These statutory retirement ages induce retirement behavior—individuals are more likely to retire if they have reached the minimum of early retirement—but they are unrelated to an individual's health. Using these variations in statutory retirement ages across European countries as instruments, Coe and Zamarro (217) found that retirement has a positive effect on overall general health, leading to a short-term decrease in the probability of reporting poor health, and a long-lasting improvement in the overall health index.

Several studies have also used this approach to examine the impact of retirement on cognitive function, a topic discussed briefly in Chapter 5, focused on working conditions, rather than policies, linked to health. Based on this approach, Rohwedder and Willis (215) used variation across European countries, United States, and England and found that early retirement has a negative effect on cognitive ability. Their findings, however, are not confirmed in other studies. Based on data from the HRS, Coe and colleagues (218) used employers' offers of early retirement windows—an incentive offered by a firm to leave at a particular time—as instruments to assess the impact of retirement on late-life cognitive function. Because employers must offer these windows without discriminating workers individually, they serve as good instruments to identify this effect. They find a negative association between time in retirement and cognitive function. However, instrumental variation estimates cast doubts on the causal nature of this association. Time in retirement was unrelated to cognitive function among white-collar workers, and it may have a positive effect on cognitive function for blue-collar workers.

Based on the HRS, Calvo and colleagues (219) used Social Security's full retirement age and unexpected early retirement window offers as instrumental variables. Interestingly, their results suggest that the effect of retirement depends on timing. Those retiring before age 62 seem to fare worse than those who continue to work; however, retirement at ages 62 and beyond is not associated with worse physical and mental health. If anything, those retiring at age 62 and beyond might fare better than those continuing to work. Because age 62 is the age at which workers can start

claiming social security benefits in the United States, retirement beyond this age is considered the "normal" age of retirement. Thus, any retirement before that age is likely to reflect specific circumstances or selection not well accounted for by the instrumental variable approach. In general, these findings again point to no negative effects of retirement on physical or mental health. Other studies have also found that retirement has no impact on mortality (220).

Studies have also explored the impact of retirement on health behavior. Changes in physical activity, dietary habits, and weight have been commonly assessed outcomes, given the direct implications of retirement for time allocations potentially influencing these behaviors. In a recent systematic review, Barnett and colleagues (221) reviewed 19 studies examining the impact of retirement on physical activity. Their results suggest that leisure physical activity increases after retirement, but there is no clear pattern for total physical activity. In particular, lower SES groups seem to experience a decrease in both leisure and total physical activity after retirement, while higher SES groups experience an increase. Sjosten and colleagues (222) used data from the GAZEL study to examine physical activity trajectories before and after retirement and found that the transition to retirement was associated with a substantial increase in leisure-time physical activity, which may have translated into less weight gain.

Based on data from the HRS, Chung and colleagues (223) used individual fixed effect models to assess whether transitions to retirement were associated with changes in total physical activity among older US workers. They found that physical activity decreased after retirement from a physically demanding job, but increased for workers in sedentary jobs. In a separate analysis of the same data, Chung and colleagues (224) found that retirement led overall to a modest increase in weight, but this effect was confined to workers already overweight, less wealthy, and in physically demanding occupations, while no effect was observed for wealthier works and those in sedentary jobs. Using the same data, Chung and colleagues (225) examined the impact of retirement on out-eating and weight change and concluded that retirement may increase the time available for food preparation at home and reduce weight.

The studies discussed above reflect the variety of approaches and perspectives on how retirement decisions influence health. Nonetheless, the hypothesis that later retirement is beneficial to health does not seem to come out strongly either in longitudinal studies or in most studies that explicitly account for endogenous retirement decisions using firm-specific early retirement window offers or country-specific statutory retirement age policies. If anything, it would seem as if retirement leads to a short-term improvement in mental health, with no clear effects on physical health outcomes immediately after retirement. Some changes after retirement are also evident in health-related behavior, with some studies suggesting that retirement might decrease the frequency of eating out and might increase leisure physical activity.

What do these results tell us about the potential impact of recently enacted policies to increase retirement age for future generations across many countries? On the one hand, the evidence would seem to suggest that extending retirement age will lead to poorer mental health, as those who retire seem to exhibit better mental health outcomes than those who continue to work. On the other hand, although results are still inconclusive, it would seem as if increasing retirement age could worsen mental health, but there seems to be little evidence that it would harm physical health. There is a gap in knowledge, however, about the long-run effects of retirement. Most studies have only examined short-term changes in health after retirement; whether retirement influences long-run physical and mental health trajectories, as well as survival, remains an open question.

A final important consideration refers to the differential effect of retirement across countries with different institutions, and prospects for older people. For example, the studies that have found that retirement harms health come primarily from the United States, while most studies in European countries have found little evidence that retirement has negative health effects. This may reflect profound differences between the United States and European countries in the economic and social well-being of older people, their level of social participation, volunteering, family contacts, and the extent of social networks after retirement. Further research is required to disentangle whether institutional and cultural features associated with the level of engagement after retirement might explain the discrepancy of findings on the impact of retirement on health across countries.

CONCLUSION AND FUTURE DIRECTIONS

We have reviewed a large body of evidence documenting the challenges of establishing whether job loss, maternity leave, and retirement causally influence health. A key finding from studies over the last decade is that the direction between employment and health is bidirectional: The impact of poor health on the ability to work—a link underexamined by social epidemiologists—seems to be an important reason for the finding that employed workers have better health than unemployed workers and those out of the labor market. Quasi-experimental studies over the last years suggest that unemployment may have an impact on health, but findings are inconsistent, and the effect appears more solid for mental health than for physical health outcomes.

The emphasis of research over the last decades on unemployment has drawn the attention away from other characteristics of employment that may be equally important for the health of current and future generations. We have discussed an emerging body of evidence suggesting that policies to support working parents with children, as well as older workers, may also have important health effects. A focus on labor policies as point of departure has two potential advantages: First, studying the impact of employment policies will offer insights into how future policy reforms and institutions may influence health. Second, variations in policy across countries and over time offer a unique opportunity to identify the causal impact of employment on health, as they often come as exogenous shocks that change the dynamics of employment. Social epidemiologists, so far, have been not fully exploited these opportunities to understand the impact of policies. We emphasize the need for a shift in focus toward studies that directly assess the impact of specific policies and institutions on health.

A focus on policies would be a welcome step forward for research on the impact of job insecurity and underemployment on health. While an appealing hypothesis, current research has struggled to disentangle whether these aspects of employment causally influence health. Over the last years, an interest has emerged in how legislation that loosened the requirements for hiring temporary workers, which increased the incidence of fixed-term contracts, influences incidence of work injuries and health (166, 226, 227). This literature is motivated by claims that the type of contract a worker has—namely temporary or permanent—has a causal effect on the risk of work injuries. Again, this idea is based on the incentives for investing in human capital provided by social protection policies. Employers have fewer incentives to invest in human capital for temporary short-term contracts than for identical workers in permanent contracts (228) because the returns are smaller

for the former. This would lead to lower skills and expertise in temporary workers, as well as lower investments in safety, resulting in higher accident rates. A promising approach would be to exploit changes in legislation to understand how contractual arrangements influence health. Over the last decades, many European governments have taken measures to reduce labor market rigidities, such as lowering firing costs, but the health impact of these measures remains poorly understood.

Modern welfare states increasingly conceive of social protection as a means to promote human capital. Employment protection policies, for example, may incentivize worker and employer investments in training and safety, and may through these mechanisms improve health. Maternity leave policies may increase job continuity and help mothers retain use of skills, diminishing the long-term negative effects of fertility decisions. Maternity leave policies may thus lead to better health in the long run among women by improving women's careers and earning trajectories. Research on these mechanisms requires a shift from a focus on the short-term effects of employment on health, toward a focus on how institutions shape the long-term career and income trajectories of workers, and how this ultimately influences their health in the long run. Under current retirement laws, a large majority of current generations of workers will spend several years or decades in retirement. A key question is thus how to create the institutions and policies that will generate a healthy working force as well as an active and engaged older population in countries currently facing major demographic shifts.

The large discrepancy in findings from studies across different countries suggests that national institutions and policies shape whether and how employment impacts health. For example, job loss might have weaker health effects in the Nordic Scandinavian countries, where unemployment income and other benefits are very accessible, as compared with the United States, where unemployment provisions are limited and restricted. On the other hand, very generous unemployment benefits may also increase the duration of unemployment spells and diminish human capital accumulation, ultimately harming long-term health. A research agenda that focuses on disentangling how these aspects of national employment legislation influence health will be essential in advancing our understanding of the health effects of employment. Many of these regulations are enacted at the national level; as a result, cross-national comparative studies might often be the only way to understand the impact of national policies on population health. The large variability in employment policy reforms, and the different policy responses to macroeconomic shocks over the last decades, offers a unique opportunity for social epidemiologists to study how employment policies influence health.

Several key societal challenges suggest that focusing on policies that affect particular groups will be essential. Employment policies that affect mothers and their subsequent careers will be essential. In addition, as women have entered the labor force and cultural norms have changed, fathers have become increasingly involved in childcare activities; fathers, too, face the challenges of combining work and family responsibilities. Will these changes have any effects on the future health of fathers? How do policies that promote fathers' involvement affect their health, as well as the health of mothers and their children? As we have seen, research on family support policies has traditionally focused on maternity leave policies around the period of birth. However, a potential, more enduring, challenge in combining work and family responsibilities comes during the years following childbirth. How do policies that support parents in the early years of child development impact both parental health and that of their children? There is great variation in these policies across countries. For example, childcare support policies differ dramatically across Finland and

the United States. Since 1973, Finnish policy facilitates full-time employment among women by providing families with young children guaranteed access to subsidized childcare over the preschool years (229, 230). In contrast, the United States lacks any universal model of subsidized care, with programs such as Head Start targeted to families already in great socioeconomic disadvantage. In Norway, parents are entitled to a period of shared paternity leave in the period after childbirth. Whether and how these policies affect both fathers' and mothers' health, as well as the health of their children, is yet to be established.

In coming decades, flexible work arrangements are likely to become the norm rather than the exception. With demographic transitions related to both aging and women's participation in the labor force, the need for taking care of not only children but also aging parents, spouses, and family members will grow. A key challenge will therefore be to establish how changes in the flexibility of work arrangements impact the health of parents and their children and other family members. In the UK, 63% of mothers and 89% of fathers aged 25–34 and 75% of mothers and 92% of fathers aged 35–49 are in the labor force (231). Difficulties in balancing the competing needs of work and family life are therefore a concern for many working families (4). In 2002, under the Flexible Working Act (232), UK parents with children under the age of 6 have the right to request flexible work arrangements (233). The share of workplaces offering at least one form of flexible work arrangement increased from under a quarter in 1998 to more than 90% in 2008 (234, 235). Between 2003 and 2006, increased availability was noted for flexi-time (48% in 2003 vs. 53% in 2006), job-share (41% vs. 47%), and term-time working (32% vs. 37%) and compressed working week arrangements (30% vs. 35%) (236). The proportion of employees using at least one flexible working practice in the previous year rose from 51% in 2003 to 56% in 2006 (236). From 2006 to 2011, availability of teleworking increased from 14% to 59%, and the possibility of a career break rose from 29% to 46%. These changes in the flexibility of work arrangements could potentially have both negative and positive spillover effects on the health of parents and their children.

Equally important is the focus on older populations who will face the prospect of longer lives as well as prolonged working careers, and how this will affect their health at old age. Most studies have focused on policies that regulate the age of retirement; expanding this focus to policies that influence the active participation of older populations after retirement will also be essential. For example, active involvement of older populations in the care of grandchildren may influence their health and well-being, as well as that of young working families. Incentives to participate in volunteering activities may bring benefits to the health of some older people, but it may also harm those who derive less utility from these roles. Much research in economics has focused on the labor market effects of disability insurance policies, which often serve as pathways to retirement for older workers in some European countries. Beyond labor market effects, these policies may also influence the health of older workers in the pre- and post-retirement years. Unemployment benefits targeted for older workers may also have important effects not only on labor market decisions but also on the health of older workers and their families.

Policies that affect the labor market trajectories of young workers or school graduates in the transition to the labor market may be essential. As we have shown, there is evidence that the early years of a young worker's career can have profound effects on their long-term career and earnings trajectories, which may ultimately influence their health in later life. How do policies that shape the employment opportunities of young workers influence their long-run health? Again, the impact of these policies in human capital accumulation is key to understanding these issues.

Active labor market programs, for example, may provide young people the opportunity to invest in human capital as they make a rough transition to the labor market during an economic downturn. In turn, this may ease the effect of harsh economic conditions on their career trajectories, and ultimately improve their current and late-life health.

In conclusion, we have seen major changes in the nature of work over the last years, and understanding how these changes will impact the health of future generations will be the key challenges for social epidemiologists trying to disentangle the impact of work on health. A focus on policies and the large variability in institutions across different countries and over time offers a unique opportunity to study the causal impact of these changes on health. The health impact of policies across different domains, including those shaping job security, work flexibility and parental leave, young workers' careers and training, and policies shaping the transition to retirement, might have important health effects that will require our attention over the next decades. A shift toward focusing on approaches that enable us to determine whether the effects of these policies is causal requires a revision of current epidemiological approaches and the integration of econometric and other methodological techniques that focus on causality. Social protection policies as a means to promote human capital formation offer a useful framework that can guide our theoretical understanding of how employment policies impact the health of populations. Focusing on cross-national studies will enable us to establish how national employment policies impact population health, and how populations across different countries respond differently to these policies.

REFERENCES

1. Brugiavini A, Pasini G, Trevisan E. The direct impact of maternity benefits on leave taking: evidence from complete fertility histories. Adv Life Course Res. 2012;18(1):46–67.
2. Rossin M. The effects of maternity leave on children's birth and infant health outcomes in the United States. J Health Econ. 2011;30(2):221–39.
3. Rossin-Slater M, Ruhm CJ, Waldfogel J. The effects of California's paid family leave program on mothers' leave-taking and subsequent labor market outcomes. J Policy Anal Manage. 2013;32(2):224–45.
4. Ruhm CJ. Policies to assist parents with young children. Future Child. 2011;21(2):37–68.
5. Klerman JA, Leibowitz A. Labor supply effects of state maternity leave legislation. In: Blau FD, Ehrenberg RG, editors. Gender and family issues in the workplace. New York: Russell Sage; 2000.
6. Grossman M. On the concept of health capital and the demand for health. J Polit Econ. 1972;80(2):223–55.
7. Galama T, Kapteyn A. Grossman's missing health threshold. J Health Econ. 2011;30(5):1044–56.
8. Galama T, Kapteyn A, Fonseca R, Michaud PC. A health production model with endogenous retirement. Health Econ. 2012;22(8):883–902.
9. Galama TJ, Hullegie P, Meijer E, Outcault S. Is there empirical evidence for decreasing returns to scale in a health capital model? Health Econ. 2012;21(9):1080–100.
10. Currie J, Madrian B. Health, health insurance and the labor market. In: Ashenfelter O, Card D, editors. Handbook of labor economics. Amsterdam: Elsevier Science; 1999.
11. Gordo L. Effects of short- and long-term unemployment on health satisfaction: evidence from German data. Appl Econ. 2006;38:2335–50.
12. Klerman JA, Leibowitz A. Labor supply effects of state maternity leave legislation. New York: Russell Sage Foundation; 1997.

13. Waldfogel J, Washbrook E. Early years policy: child development research. 2011; 2011.
14. Ruhm CJ. Parental leave and child health. J Health Econ. 2000;19(6):931–60.
15. Staehelin K, Bertea PC, Stutz EZ. Length of maternity leave and health of mother and child—a review. Int J Public Health. 2007;52(4):202–9.
16. Tanaka S. Parental leave and child health across OECD countries. Econ J. 2005;115(501):F7–28.
17. Janlert U, Hammarstrom A. Which theory is best? Explanatory models of the relationship between unemployment and health. BMC Public Health. 2009;9:235.
18. Bartley M. Unemployment and ill health: understanding the relationship. J Epidemiol Commun H. 1994;48(4):333–7.
19. Jackson PR, Warr PB. Unemployment and psychological ill-health: the moderating role of duration and age. Psychol Med. 1984;14(3):605–14.
20. Leeflang RL, Klein-Hesselink DJ, Spruit IP. Health effects of unemployment—II. Men and women. Soc Sci Med. 1992;34(4):351–63.
21. Mattiasson I, Lindgarde F, Nilsson JA, Theorell T. Threat of unemployment and cardiovascular risk factors: longitudinal study of quality of sleep and serum cholesterol concentrations in men threatened with redundancy. BMJ. 1990;301(6750):461–6.
22. Martikainen P, Maki N, Jantti M. The effects of unemployment on mortality following workplace downsizing and workplace closure: a register-based follow-up study of Finnish men and women during economic boom and recession. Am J Epidemiol. 2007;165(9):1070–5.
23. Martikainen PT. Unemployment and mortality among Finnish men, 1981–5. BMJ. 1990;301(6749):407–11.
24. Ahs AM, Westerling R. Mortality in relation to employment status during different levels of unemployment. Scand J Public Health. 2006;34(2):159–67.
25. Eliason M, Storrie D. Job loss is bad for your health—Swedish evidence on cause-specific hospitalization following involuntary job loss. Soc Sci Med. 2009;68(8):1396–406.
26. Schmieder JF, von Wachter T, Bender S. The effects of extended unemployment insurance over the business cycle: evidence from regression discontinuity estimates over 20 years. Q J Econ. 2012;127(2):701–52.
27. Lalive R, Schlosser A, Steinhauer A, Zweimüller J. Parental leave and mothers' careers: the relative importance of job protection and cash benefits. Review of Economic Studies. 10.1093/restud/rdt028 2013;
28. Jahoda M. Employment and unemployment: A social psychological analysis. New York: Press Syndicate of the University of Cambridge; 1982.
29. Warr PB. Work, unemployment, and mental health. Oxford New York: Clarendon Press; Oxford University Press; 1987.
30. Vinokur AD, van Ryn M, Gramlich EM, Price RH. Long-term follow-up and benefit-cost analysis of the Jobs Program: a preventive intervention for the unemployed. J Appl Psychol. 1991;76(2):213–9.
31. Hintikka J, Lehto SM, Niskanen L, Huotari A, Herzig KH, Koivumaa-Honkanen H, et al. Unemployment and ill health: a connection through inflammation? BMC Public Health. 2009;9:410.
32. Dooley CD, Prause J. The social costs of unemployment: inadequate employment as disguised unemployment. Cambridge: Cambridge University Press; 2004.
33. Heckhausen J, Schulz R. A Life-span theory of control. Psychol Rev. 1995;102:284–304.
34. Karasek R. Healthy work: stress, productivity, and the reconstruction of working life. New York: Basic Books; 1990.
35. Morris JK, Cook DG. A critical review of the effect of factory closures on health. Br J Ind Med. 1991;48(1):1–8.
36. Ruhm CJ. Economic conditions and alcohol problems. J Health Econ. 1995;14(5):583–603.
37. Ruhm CJ. Healthy living in hard times. J Health Econ. 2005;24(2):341–63.

38. Ellwood DT. Teenage unemployment: permanent scars or temporary blemishes? In: Freeman RB, Wise DA, editors. The youth labor market problem: its nature, causes, and consequences. Chicago: University of Chicago Press; 1982. pp. 349–84.

39. Ruhm CJ. Are workers permanently scarred by job displacements? Am Econ Rev. 1991;81(1):319–24.

40. Arulampalam W. Is unemployment really scarring? effects of unemployment experiences on wages. Econ J. 2001;111(475):F585–606.

41. Gangl M. Welfare states and the scar effects of unemployment: a comparative analysis of the United States and West Germany. Am J Sociol. 2004;109(6):1319–64.

42. Gangl M. Scar effects of unemployment: an assessment of institutional complementarities. Am Sociol Rev. 2006;71(6):986–1013.

43. Knabe A, Ratzel S. Scarring or scaring? The psychological impact of past unemployment and future unemployment risk. Economica. 2011;78(310):283–93.

44. Dooley D. Unemployment, underemployment, and mental health: conceptualizing employment status as a continuum. Am J Community Psychol. 2003;32(1–2):9–20.

45. Smith R. Unemployment and health: a disaster and a challenge. Oxford: Oxford University Press; 1987.

46. Jin RL, Shah CP, Svoboda TJ. The impact of unemployment on health: a review of the evidence. CMAJ. 1995;153(5):529–40.

47. Dooley D, Fielding J, Levi L. Health and unemployment. Annu Rev Public Health. 1996;17:449–65.

48. Burgard SA, Brand JE, House JS. Toward a better estimation of the effect of job loss on health. J Health Soc Behav. 2007;48(4):369–84.

49. Schroder M. Scar or Blemish? Investigating the long-term impact of involuntary job loss on health. In: Borsch-Supan A, Brandt M, Hank K, Schroder M, editors. The individual and the welfare state: life histories in Europe. New York and Heidelberg: Springer; 2011. pp. 191–201.

50. Bartley M, Sacker A, Clarke P. Employment status, employment conditions, and limiting illness: prospective evidence from the British household panel survey 1991–2001. J Epidemiol Community Health. 2004;58(6):501–6.

51. Booker CL, Sacker A. Psychological well-being and reactions to multiple unemployment events: adaptation or sensitisation? J Epidemiol Community Health. 2012;66(9):832–8.

52. Gallo WT, Teng HM, Falba TA, Kasl SV, Krumholz HM, Bradley EH. The impact of late career job loss on myocardial infarction and stroke: a 10 year follow up using the health and retirement survey. Occup Environ Med. 2006;63(10):683–7.

53. Gallo WT, Bradley EH, Falba TA, Dubin JA, Cramer LD, Bogardus ST Jr, et al. Involuntary job loss as a risk factor for subsequent myocardial infarction and stroke: findings from the Health and Retirement Survey. Am J Ind Med. 2004;45(5):408–16.

54. Gallo WT, Bradley EH, Siegel M, Kasl SV. Health effects of involuntary job loss among older workers: findings from the health and retirement survey. J Gerontol B Psychol Sci Soc Sci. 2000;55(3):S131–40.

55. Gallo WT, Bradley EH, Siegel M, Kasl SV. The impact of involuntary job loss on subsequent alcohol consumption by older workers: findings from the health and retirement survey. J Gerontol B Psychol Sci Soc Sci. 2001;56(1):S3–9.

56. Falba T, Teng HM, Sindelar JL, Gallo WT. The effect of involuntary job loss on smoking intensity and relapse. Addiction. 2005;100(9):1330–9.

57. Gallo WT, Bradley EH, Dubin JA, Jones RN, Falba TA, Teng HM, et al. The persistence of depressive symptoms in older workers who experience involuntary job loss: results from the health and retirement survey. J Gerontol B Psychol Sci Soc Sci. 2006;61(4):S221–8.

58. Berchick ER, Gallo WT, Maralani V, Kasl SV. Inequality and the association between involuntary job loss and depressive symptoms. Soc Sci Med. 2012;75(10):1891–4.

59. Strully K. Job loss and health in the US labor market. Demography. 2009;46(2):221–46.

60. Salm M. Does job loss cause ill health? Health Econ. 2009;18(9):1075–89.
61. Gallo WT, Bradley EH, Teng HM, Kasl SV. The effect of recurrent involuntary job loss on the depressive symptoms of older US workers. Int Arch Occup Environ Health. 2006;80(2):109–16.
62. Gallo WT, Brand JE, Teng HM, Leo-Summers L, Byers AL. Differential impact of involuntary job loss on physical disability among older workers does predisposition matter? Res Aging. 2009;31(3):345–60.
63. Bockerman P, Ilmakunnas P. Unemployment and self-assessed health: evidence from panel data. Health Econ. 2009;18(2):161–79.
64. Browning M, Dano AM, Heinesen E. Job displacement and stress-related health outcomes. Health Econ. 2006;15(10):1061–75.
65. Schmitz H. Why are the unemployed in worse health? The causal effect of unemployment on health. Labour Econ. 2011;18(1):71–8.
66. Romeu Gordo L. Effects of short- and long-term unemployment on health satisfaction: evidence from German data. Appl Econ. 2006;38(20):2335–50.
67. Kaufman JS. Commentary: why are we biased against bias? Int J Epidemiol. 2008;37(3):624–6.
68. McLeod CB, Lavis JN, Macnab YC, Hertzman C. Unemployment and mortality: a comparative study of Germany and the United States. Am J Public Health. 2012;102(8):1542–50.
69. Glymour MM. Sensitive periods and first difference models: integrating etiologic thinking into econometric techniques: a commentary on Clarkwest's "Neo-materialist theory and the temporal relationship between income inequality and longevity change". Soc Sci Med. 2008;66(9):1895–902; discussion 903–8.
70. Moser KA, Goldblatt PO, Fox AJ, Jones DR. Unemployment and mortality: comparison of the 1971 and 1981 longitudinal study census samples. Br Med J (Clin Res Ed). 1987;294(6564):86–90.
71. Moser KA, Fox AJ, Jones DR, Goldblatt PO. Unemployment and mortality: further evidence from the OPCS Longitudinal Study 1971–81. Lancet. 1986;1(8477):365–7.
72. Moser KA, Fox AJ, Jones DR. Unemployment and mortality in the OPCS Longitudinal Study. Lancet. 1984;2(8415):1324–9.
73. Roelfs DJ, Shor E, Davidson KW, Schwartz JE. Losing life and livelihood: a systematic review and meta-analysis of unemployment and all-cause mortality. Soc Sci Med. 2011;72(6):840–54.
74. Iversen L, Andersen O, Andersen PK, Christoffersen K, Keiding N. Unemployment and mortality in Denmark, 1970–80. Br Med J (Clin Res Ed). 1987;295(6603):879–84.
75. Rogot E, Sorlie PD, Johnson NJ. Life expectancy by employment status, income, and education in the National Longitudinal Mortality Study. Public Health Rep. 1992;107(4):457–61.
76. Sorlie PD, Rogot E. Mortality by employment status in the National Longitudinal Mortality Study. Am J Epidemiol. 1990;132(5):983–92.
77. Morris JK, Cook DG, Shaper AG. Loss of employment and mortality. BMJ. 1994;308(6937):1135–9.
78. Bethune A. Economic activity and mortality of the 1981 Census cohort in the OPCS longitudinal study. Popul Trends. 1996 Spring(83):37–42.
79. Martikainen PT, Valkonen T. Excess mortality of unemployed men and women during a period of rapidly increasing unemployment. Lancet. 1996;348(9032):909–12.
80. Maki N, Martikainen P. A register-based study on excess suicide mortality among unemployed men and women during different levels of unemployment in Finland. J Epidemiol Community Health. 2012;66(4):302–7.
81. Sullivan D, von Wachter T. Job displacement and mortality: an analysis using administrative data. Q J Econ. 2009;124(3):1265–306.
82. Jacobson L, LaLonde R, Sullivan D. Earnings losses of displaced workers. Am Econ Rev. 1993;83:685–709.
83. Eliason M, Storrie D. Does job loss shorten life? J Hum Resour. 2009;44(2):277–302.

84. Browning M, Heinesen E. Effect of job loss due to plant closure on mortality and hospitalization. J Health Econ. 2012;31(4):599–616.
85. Ogburn W, Thomas D. The influence of the business cycle on certain social conditions. JAMA. 1922;18(139):324–40.
86. Tapia Granados JA. Increasing mortality during the expansions of the US economy, 1900–1996. Int J Epidemiol. 2005;34(6):1194–202.
87. Thomas D. Social aspects of the business cycle. London: Routledge; 1925.
88. Brenner MH. Economic changes and heart disease mortality. Am J Public Health. 1971;61(3):606–11.
89. Brenner MH. Mortality and the national economy: a review, and the experience of England and Wales, 1936–76. Lancet. 1979;2(8142):568–73.
90. Brenner MH. Unemployment, economic growth, and mortality. Lancet. 1979;1(8117):672.
91. Brenner MH. Economic indicators as predictors of ill-health. Lancet. 1981;2(8240):262.
92. Gravelle HS. Time series analysis of mortality and unemployment. J Health Econ. 1984;3(3):297–305.
93. Gravelle HS, Hutchinson G, Stern J. Mortality and unemployment: a critique of Brenner's time-series analysis. Lancet. 1981;2(8248):675–9.
94. Wagstaff A. Time series analysis of the relationship between unemployment and mortality: a survey of econometric critiques and replications of Brenner's studies. Soc Sci Med. 1985;21(9):985–96.
95. Ruhm CJ. Are recessions good for your health? Q J Econ. 2000;115(2):617–50.
96. Ruhm CJ. Good times make you sick. J Health Econ. 2003;22(4):637–58.
97. Ruhm CJ. Commentary: mortality increases during economic upturns. Int J Epidemiol. 2005;34(6):1206–11.
98. Ruhm CJ. A healthy economy can break your heart. Demography. 2007;44(4):829–48.
99. Ruhm CJ, Black WE. Does drinking really decrease in bad times? J Health Econ. 2002;21(4):659–78.
100. Tapia Granados JA, Diez Roux AV. Life and death during the Great Depression. Proc Natl Acad Sci U S A. 2009;106(41):17290–5.
101. Ariizumi H, Schirle T. Are recessions really good for your health? Evidence from Canada. Soc Sci Med. 2012;74(8):1224–31.
102. Gerdtham UG, Ruhm CJ. Deaths rise in good economic times: evidence from the OECD. Econ Hum Biol. 2006;4(3):298–316.
103. Ryan E. Who is hurt by procyclical mortality? Soc Sci Med. 2008;67(12):2051–8.
104. Tapia Granados JA, Ionides EL. The reversal of the relation between economic growth and health progress: Sweden in the 19th and 20th centuries. J Health Econ. 2008 May;27(3):544–63.
105. Miller DL, Page ME, Stevens AH, Filipsky M. Why are recessions good for your health? Am Econ Rev: Papers and Proceedings. 2009;99(2):122–7.
106. Neumayer E. Recessions lower (some) mortality rates: evidence from Germany. Soc Sci Med. 2004;58(6):1037–47.
107. Neumayer E. Commentary: the economic business cycle and mortality. Int J Epidemiol. 2005;34(6):1221–2.
108. Dehejia R, Lleras-Muney A. Booms, busts, and babies' health. Q J Econ. 2004;119(3):1091–130.
109. Catalano R, Goldman-Mellor S, Saxton K, Margerison-Zilko C, Subbaraman M, LeWinn K, et al. The health effects of economic decline. Annu Rev Public Health. 2011;32:431–50.
110. Elder G. Children of the great depression. Boulder, CO: Westview; 1974.
111. Elder G. Children of the land. Chicago: University of Chicago Press; 2000.
112. Lee D, Brooks-Gunn J, McLanahan SS, Notterman D, Garfinkel I. The Great Recession, genetic sensitivity, and maternal harsh parenting. Proc Natl Acad Sci U S A. 2013;110(34):13780–4.

113. Loewenstein GF, Weber EU, Hsee CK, Welch N. Risk as feelings. Psychol Bull. 2001;127(2):267–86.

114. Baumeister RF, Vohs KD, DeWall CN, Zhang L. How emotion shapes behavior: feedback, anticipation, and reflection, rather than direct causation. Pers Soc Psychol Rev. 2007;11(2):167–203.

115. Gerdtham UG, Johannesson M. Business cycles and mortality: results from Swedish microdata. Soc Sci Med. 2005;60(1):205–18.

116. Svensson M. Do not go breaking your heart: do economic upturns really increase heart attack mortality? Soc Sci Med. 2007;65(4):833–41.

117. Ruhm C. Recessions healthy no more. NBER Working Paper No 19287. 2013; August.

118. Tapia Granados JA. Economic growth and health progress in England and Wales: 160 years of a changing relation. Soc Sci Med. 2012;74(5):688–95.

119. Gonzalez F, Quast T. Mortality and business cycles by level of development: evidence from Mexico. Soc Sci Med. 2010;71(12):2066–73.

120. Gonzalez F, Quast T. Macroeconomic changes and mortality in Mexico. Empir Econ. 2011;40(2):305–19.

121. Bhalotra S. Fatal Fluctuations? Cyclicality in infant mortality in India. J Dev Econ. 2010;93(1):7–19.

122. Suhrcke M, Stuckler D. Will the recession be bad for our health? It depends. Soc Sci Med. 2012;74(5):647–53.

123. Yoon JH, Junger W, Kim BW, Kim YJ, Koh SB. Investigating the time lag effect between economic recession and suicide rates in agriculture, fisheries, and forestry workers in Korea. Saf Health Work. 2012;3(4):294–7.

124. Reeves A, Stuckler D, McKee M, Gunnell D, Chang SS, Basu S. Increase in state suicide rates in the USA during economic recession. Lancet. 2012;380(9856):1813–4.

125. Nandi A, Prescott MR, Cerda M, Vlahov D, Tardiff KJ, Galea S. Economic conditions and suicide rates in New York City. Am J Epidemiol. 2012;175(6):527–35.

126. Barr B, Taylor-Robinson D, Scott-Samuel A, McKee M, Stuckler D. Suicides associated with the 2008–10 economic recession in England: time trend analysis. BMJ. 2012;345:e5142.

127. Luo F, Florence CS, Quispe-Agnoli M, Ouyang L, Crosby AE. Impact of business cycles on US suicide rates, 1928–2007. Am J Public Health. 2011;101(6):1139–46.

128. Blasco-Fontecilla H, Perez-Rodriguez MM, Garcia-Nieto R, Fernandez-Navarro P, Galfalvy H, de Leon J, et al. Worldwide impact of economic cycles on suicide trends over 3 decades: differences according to level of development. A mixed effect model study. BMJ Open. 2012;2(3).

129. Chen VC, Chou JY, Lai TJ, Lee CT. Suicide and unemployment rate in Taiwan, a population-based study, 1978–2006. Soc Psychiatry Psychiatr Epidemiol. 2010;45(4):447–52.

130. Saurina C, Bragulat B, Saez M, Lopez-Casasnovas G. A conditional model for estimating the increase in suicides associated with the 2008–2010 economic recession in England. J Epidemiol Community Health. 2013;67(9):779–87.

131. Davalos ME, French MT. This recession is wearing me out! Health-related quality of life and economic downturns. J Ment Health Policy Econ. 2011;14(2):61–72.

132. Gili M, Roca M, Basu S, McKee M, Stuckler D. The mental health risks of economic crisis in Spain: evidence from primary care centres, 2006 and 2010. Eur J Public Health. 2013;23(1):103–8.

133. Xu X. The business cycle and health behaviors. Soc Sci Med. 2013;77:126–36.

134. Bockerman P, Johansson E, Helakorpi S, Prattala R, Vartiainen E, Uutela A. Does a slump really make you thinner? Finnish micro-level evidence 1978–2002. Health Econ. 2007;16(1):103–7.

135. Latif E. The impact of macroeconomic conditions on obesity in Canada. Health Econ. 2013 Jul 3.

136. Dave DM, Kelly IR. How does the business cycle affect eating habits? Soc Sci Med. 2012;74(2):254–62.

137. Dee TS. Alcohol abuse and economic conditions: evidence from repeated cross-sections of individual-level data. Health Econ. 2001;10(3):257–70.

138. Nandi A, Charters TJ, Strumpf EC, Heymann J, Harper S. Economic conditions and health behaviours during the "Great Recession." J Epidemiol Community Health. 2013 Aug 22.

139. Johansson E, Bockerman P, Prattala R, Uutela A. Alcohol-related mortality, drinking behavior, and business cycles: are slumps really dry seasons? Eur J Health Econ. 2006;7(3):215–20.

140. Lo CC, Cheng TC. Heavy drinking during periods of high unemployment: 15-year trend study of the role of race/ethnicity. Drug Alcohol Depend. 2013 Jul 20.

141. Hoynes HW, Miller DL, Schaller J. Who suffers during recessions? NBER Working Paper 17951. 2012.

142. Fontenla M, Gonzalez F, Quast T. Are recessions good for everyone's health? The association between mortality and the business cycle by race/ethnicity in the US. Appl Econ Lett. 2011;18(3):207–12.

143. Catalano R. The health effects of economic insecurity. Am J Public Health. 1991;81(9):1148–52.

144. Kivimäki M, Honkonen T, Wahlbeck K, Elovainio M, Pentti J, Klaukka T, et al. Organisational downsizing and increased use of psychotropic drugs among employees who remain in employment. J Epidemiol Community Health. 2007;61(2):154–8.

145. Kivimäki M, Vahtera J, Pentti J, Ferrie JE. Factors underlying the effect of organisational downsizing on health of employees: longitudinal cohort study. BMJ. 2000;320(7240):971–5.

146. Wiemers E. The effect of unemployment on household composition and doubling up. Working Paper: National Poverty Center. 2010.

147. Mykyta L, Macartney S. The effects of recession on household composition: "doubling up" and economic well-being. US Census Bureau. SEHSD Working Paper Number 2011-4 2011.

148. van den Berg GJ, Doblhammer G, Christensen K. Exogenous determinants of early-life conditions, and mortality later in life. Soc Sci Med. 2009;68(9):1591–8.

149. van den Berg GJ, Doblhammer-Reiter G, Christensen K. Being born under adverse economic conditions leads to a higher cardiovascular mortality rate later in life: evidence based on individuals born at different stages of the business cycle. Demography. 2011;48(2):507–30.

150. Barker DJP. Mothers, babies, and health in later life. 2nd ed. Edinburgh; New York: Churchill Livingstone; 1998.

151. Maclean JC. The health effects of leaving school in a bad economy. J Health Econ. 2013;32(5):951–64.

152. Hessel P, Avendano M. Are economic recessions at the time of leaving school associated with worse physical functioning in later life? Ann Epidemiol. 2013; forthcoming (in press).

153. Coile CC, Levine PB. Recessions, retirement, and social security. Am Econ Rev. 2011;101(3):23–8.

154. Coile CC, Levine PB. Labor market shocks and retirement: do government programs matter? J Public Econ. 2007;91(10):1902–19.

155. McInerney M, Mellor JM. Recessions and seniors' health, health behaviors, and healthcare use: analysis of the Medicare Current Beneficiary Survey. J Health Econ. 2012;31(5):744–51.

156. Coile C, Levine P, McKnight R. Recessions, older workers, and longevity: how long are recessions good for your health? Cambridge: National Bureau of Economic Research; 2012.

157. Daly M, Delaney L. The scarring effect of unemployment throughout adulthood on psychological distress at age 50: estimates controlling for early adulthood distress and childhood psychological factors. Soc Sci Med. 2013;80(0):19–23.

158. Li ZY, Page A, Martin G, Taylor R. Attributable risk of psychiatric and socio-economic factors for suicide from individual-level, population-based studies: a systematic review. Soc Sci Med. 2011;72(4):608–16.

159. Taylor R, Page A, Morrell S, Harrison J, Carter G. Mental health and socio-economic variations in Australian suicide. Soc Sci Med. 2005;61(7):1551–9.

160. Zhang J, Mckeown RE, Hussey JR, Thompson SJ, Woods JR. Gender differences in risk factors for attempted suicide among young adults: findings from the Third National Health and Nutrition Examination Survey. Ann Epidemiol. 2005;15(2):167–74.

161. Coile C, Levine P. Recessions, retirement, and social security. Gerontologist. 2011;51:437–8.

162. Coile CC, Levine PB. (2010). Implications for Retiree Well-Being. In C.C. Coile, & P.B. Levine (Eds.), Reconsidering Retirement: How Losses and Layoffs Affect Older Workers. pp. 99–116. Washington, D.C.: The Brookings Institution.

163. Coile C, Courtney C, Levine PB, McKnight R. Recessions, older workers, and longevity: how long are recessions good for your health? Am Econ J-Econ Polic. In press.

164. Quinlan M, Bohle P. Overstretched and unreciprocated commitment: reviewing research on the occupational health and safety effects of downsizing and job insecurity. Int J Health Serv. 2009;39(1):1–44.

165. Smith V. New forms of work organization. Annu Rev Sociol. 1997;23:315–39.

166. Gash V, Mertens A, Gordo L. Are fixed-term jobs bad for your health? A comparison of western Germany and Spain. Eur Soc. 2007;9(3):429–58(30).

167. Bartley M. Job insecurity and its effect on health. J Epidemiol Community Health. 2005;59(9):718–9.

168. Mohren DC, Swaen GM, van Amelsvoort LG, Borm PJ, Galama JM. Job insecurity as a risk factor for common infections and health complaints. J Occup Environ Med. 2003;45(2):123–9.

169. Winefield AH, Tiggemann M, Goldney RD. Psychological concomitants of satisfactory employment and unemployment in young people. Soc Psychiatry Psychiatr Epidemiol. 1988;23(3):149–57.

170. Winefield AH, Tiggemann M, Winefield HR. The psychological impact of unemployment and unsatisfactory employment in young men and women: longitudinal and cross-sectional data. Br J Psychol. 1991;82 (Pt 4):473–86.

171. Quinlan M, Mayhew C, Bohle P. The global expansion of precarious employment, work disorganization, and consequences for occupational health: placing the debate in a comparative historical context. Int J Health Serv. 2001;31(3):507–36.

172. Quinlan M, Mayhew C, Bohle P. The global expansion of precarious employment, work disorganization, and consequences for occupational health: a review of recent research. Int J Health Serv. 2001;31(2):335–414.

173. Rodriguez E. Marginal employment and health in Britain and Germany: does unstable employment predict health? Soc Sci Med. 2002;55(6):963–79.

174. Ferrie JE, Shipley MJ, Stansfeld SA, Marmot MG. Effects of chronic job insecurity and change in job security on self reported health, minor psychiatric morbidity, physiological measures, and health related behaviours in British civil servants: the Whitehall II study. J Epidemiol Commun H. 2002, 2002;56(6):450–4.

175. Burgard SA, Brand JE, House JS. Perceived job insecurity and worker health in the United States. Soc Sci Med. 2009;69(5):777–85.

176. Lee S, Colditz GA, Berkman LF, Kawachi I. Prospective study of job insecurity and coronary heart disease in US women. Ann Epidemiol. 2004;14(1):24–30.

177. Ferrie JE, Kivimäki M, Shipley MJ, Davey Smith G, Virtanen M. Job insecurity and incident coronary heart disease: the Whitehall II prospective cohort study. Atherosclerosis. 2013;227(1):178–81.

178. Slopen N, Glynn RJ, Buring JE, Lewis TT, Williams DR, Albert MA. Job strain, job insecurity, and incident cardiovascular disease in the Women's Health Study: results from a 10-year prospective study. PLoS ONE. 2012;7(7):e40512.

179. Laszlo KD, Engstrom K, Hallqvist J, Ahlbom A, Janszky I. Job insecurity and prognosis after myocardial infarction: The SHEEP Study. Int J Cardiol. 2012 Aug 9.

180. Virtanen P, Janlert U, Hammarstrom A. Exposure to temporary employment and job insecurity: a longitudinal study of the health effects. Occup Environ Med. 2011;68(8):570–4.

181. Laszlo KD, Pikhart H, Kopp MS, Bobak M, Pajak A, Malyutina S, et al. Job insecurity and health: a study of 16 European countries. Soc Sci Med. 2010;70(6):867–74.

182. Kalil A, Ziol-Guest KM, Hawkley LC, Cacioppo JT. Job insecurity and change over time in health among older men and women. J Gerontol B Psychol Sci Soc Sci. 2010;65B(1):81–90.

183. Rugulies R, Aust B, Burr H, Bultmann U. Job insecurity, chances on the labour market and decline in self-rated health in a representative sample of the Danish workforce. J Epidemiol Community Health. 2008;62(3):245–50.

184. Virtanen M, Nyberg ST, Batty GD, Jokela M, Heikkila K, Fransson EI, et al. Perceived job insecurity as a risk factor for incident coronary heart disease: systematic review and meta-analysis. BMJ. 2013;347:f4746.

185. Virtanen M, Kivimäki M, Joensuu M, Virtanen P, Elovainio M, Vahtera J. Temporary employment and health: a review. Int J Epidemiol. 2005;34(3):610–22.

186. Kivimäki M, Vahtera J, Virtanen M, Elovainio M, Pentti J, Ferrie JE. Temporary employment and risk of overall and cause-specific mortality. Am J Epidemiol. 2003;158(7):663–8.

187. Virtanen P, Vahtera J, Kivimäki M, Liukkonen V, Virtanen M, Ferrie J. Labor market trajectories and health: a four-year follow-up study of initially fixed-term employees. Am J Epidemiol. 2005;161(9):840–6.

188. Vahtera J, Kivimäki M, Pentti J. Effect of organisational downsizing on health of employees. Lancet. 1997;350(9085):1124–8.

189. Osthus S, Mastekaasa A. The impact of downsizing on remaining workers' sickness absence. Soc Sci Med. 2010;71(8):1455–62.

190. Osthus S. Health effects of downsizing survival and job loss in Norway. Soc Sci Med. 2012;75(5):946–53.

191. Westerlund H, Ferrie J, Hagberg J, Jeding K, Oxenstierna G, Theorell T. Workplace expansion, long-term sickness absence, and hospital admission. Lancet. 2004;363(9416):1193–7.

192. Westerlund H, Theorell T, Alfredsson L. Organizational instability and cardiovascular risk factors in white-collar employees: an analysis of correlates of structural instability of workplace organization on risk factors for coronary heart disease in a sample of 3,904 white collar employees in the Stockholm region. Eur J Public Health. 2004;14(1):37–42.

193. Ferrie JE, Westerlund H, Oxenstierna G, Theorell T. The impact of moderate and major workplace expansion and downsizing on the psychosocial and physical work environment and income in Sweden. Scand J Public Health. 2007;35(1):62–9.

194. Theorell T, Oxenstierna G, Westerlund H, Ferrie J, Hagberg J, Alfredsson L. Downsizing of staff is associated with lowered medically certified sick leave in female employees. Occup Environ Med. 2003;60(9):E9.

195. Egan M, Petticrew M, Ogilvie D, Hamilton V, Drever F. "Profits before people"? A systematic review of the health and safety impacts of privatising public utilities and industries in developed countries. J Epidemiol Commun Health. 2007;61(10):862–70.

196. Clogg CC. Measuring underemployment: demographic indicators for the United States. New York: Academic Press; 1979.

197. Sullivan T. Marginal workers, marginal jobs: University of Texas Press; 1976.

198. Robinson J. Disguised unemployment. The Economic Journal: The Quarterly Journal of the Royal Economic Society. 1936;46:225–37.

199. Friedland DS, Price RH. Underemployment: consequences for the health and well-being of workers. Am J Community Psychol. 2003;32(1–2):33–45.

200. Dooley D, Prause J, Ham-Rowbottom KA. Underemployment and depression: longitudinal relationships. J Health Soc Behav. 2000;41(4):421–36.

201. Dooley CD, Prause J. Reverse causation. In: Dooley CD, Prause J, editors. The social costs of underemployment: inadequate employment as disguised unemployment. Cambridge: Cambridge University Press; 2004. pp. 65–87.

202. US Bureau of Labor Statistics. BLS reports: women in the labor force; 2013.

203. Kramer MS, Aboud F, Mironova E, Vanilovich I, Platt RW, Matush L, et al. Breastfeeding and child cognitive development: new evidence from a large randomized trial. Arch Gen Psychiatry. 2008;65(5):578–84.

204. Baker M, Milligan K. Maternal employment, breastfeeding, and health: evidence from maternity leave mandates. J Health Econ. 2008;27(4):871–87.

205. Baker M, Milligan K. Maternal employment, breastfeeding, and health: evidence from maternity leave mandates. J Health Econ. 2008;27(4):871–87.

206. Hawkins SS, Stern AD, Gillman MW. Do state breastfeeding laws in the USA promote breast feeding? J Epidemiol Commun H. 2013;67(3):250–6.

207. Chatterji P, Markowitz S. Family leave after childbirth and the mental health of new mothers. J Ment Health Policy Econ. 2012;15(2):61–76.

208. Hyde JS, Klein MH, Essex MJ, Clark R. Maternity leave and women's mental health. Psychol Women Q. 1995;19:257–85.

209. Gjerdingen DK, Froberg DG, Kochevar L. Changes in women's mental and physical health from pregnancy through six months postpartum. J Fam Pract. 1991;32(2):161–6.

210. McGovern P, Dowd B, Gjerdingen D, Moscovice I, Kochevar L, Lohman W. Time off work and the postpartum health of employed women. Med Care. 1997;35(5):507–21.

211. Mein G, Martikainen P, Hemingway H, Stansfeld S, Marmot M. Is retirement good or bad for mental and physical health functioning? Whitehall II longitudinal study of civil servants. J Epidemiol Commun H. 2003;57(1):46–9.

212. Westerlund H, Kivimäki M, Singh-Manoux A, Melchior M, Ferrie JE, Pentti J, et al. Self-rated health before and after retirement in France (GAZEL): a cohort study. Lancet. 2009;374(9705):1889–96.

213. Jokela M, Ferrie JE, Gimeno D, Chandola T, Shipley MJ, Head J, et al. From midlife to early old age: health trajectories associated with retirement. Epidemiology. 2010;21(3):284–90.

214. Roberts BA, Fuhrer R, Marmot M, Richards M. Does retirement influence cognitive performance? the Whitehall II study. J Epidemiol Commun H. 2011;65(11):958–63.

215. Rohwedder S, Willis RJ. Mental retirement. J Econ Perspect. 2010;24(1):119–38.

216. Charles KK. Is retirement depressing? Labor force inactivity and psychological well-being in later life. In: Polachek SW, editor. Accounting for worker well-being. Research in Labor Economics, vol. 23. Amsterdam; San Diego and Oxford: Elsevier, JAI; 2004. pp. 269–99.

217. Coe NB, Zamarro G. Retirement effects on health in Europe. J Health Econ. 2011;30(1):77–86.

218. Coe NB, von Gaudecker HM, Lindeboom M, Maurer J. The effect of retirement on cognitive functioning. Health Econ. 2012;21(8):913–27.

219. Calvo E, Sarkisian N, Tamborini CR. Causal effects of retirement timing on subjective physical and emotional health. J Gerontol B Psychol Sci Soc Sci. 2013;68(1):73–84.

220. Behncke S. Does retirement trigger ill health? Health Econ. 2012;21(3):282–300.

221. Barnett I, van Sluijs EM, Ogilvie D. Physical activity and transitioning to retirement: a systematic review. Am J Prev Med. 2012;43(3):329–36.

222. Sjosten N, Kivimäki M, Singh-Manoux A, Ferrie JE, Goldberg M, Zins M, et al. Change in physical activity and weight in relation to retirement: the French GAZEL Cohort Study. BMJ Open. 2012;2:e000522.

223. Chung S, Domino ME, Stearns SC, Popkin BM. Retirement and physical activity: analyses by occupation and wealth. Am J Prev Med. 2009;36(5):422–8.

224. Chung S, Domino ME, Stearns SC. The effect of retirement on weight. J Gerontol B Psychol Sci Soc Sci. 2009;64(5):656–65.

225. Chung S, Popkin BM, Domino ME, Stearns SC. Effect of retirement on eating out and weight change: an analysis of gender differences. Obesity. 2007;15(4):1053–60.

226. Blanchard O, Landier A. The perverse effects of partial labour market reform: fixed-term contracts in france. Econ J. 2002;112(480):F214–F44.

227. Salvatori A. Labour contract regulations and workers' wellbeing: international longitudinal evidence. Labour Econ. 2010 Aug 2010;17(4):667–78.

228. Garcia-Serrano C, Hernanz V, Toharia L. Mind the gap, please! The effect of temporary help agencies on the consequences of work accidents. J Labor Res. 2010;31(2):162–82.

229. Gornick JC, Meyers MK. Families that work: policies for reconciling parenthood and employment. New York: Russell Sage Foundation Press; 2003.

230. Jacobs JA, Gerson K. The time divide: work, family and gender inequality. Cambridge, MA: Harvard University Press; 2004.

231. Office for National Statistics. Full report—women in the labour market. London: Office for National Statustics; 2013.

232. Department for Trade and Industry. Employment Act. *c 22, § 47(1–2) (UK)*. http://www.legislation.gov.uk/ukpga/2002/22/section/47 2002.

233. Department for Trade and Industry. Flexible working: the right to request and the duty to consider. London: Department for Trade and Industry; 2003.

234. Hegewisch A. Flexible working policies: a comparative review. Manchester: Equality and Human Rights Commission; 2009.

235. Confederation of British Industry. Pulling through: employment trends survey 2008. London: CBI; 2008.

236. Hooker H, Neathey F, Casebourne J, Munro M. The third Work-Life Balance Employee Survey: main findings (revised edition with corrected figures). London: Institute for Employment Studies; 2007 (ammended 2011).

CHAPTER 7

SOCIAL NETWORK EPIDEMIOLOGY

Lisa F. Berkman and Aditi Krishna

A decade ago social epidemiologists rarely incorporated the most novel social network methods into their work. It was not due to lack of knowledge about these approaches; rather, it was rare for epidemiologists to commit to the lengthy assessments involved in stronger approaches to network analysis. On the other side of this disciplinary boundary, social scientists were often relatively naïve about health assessments and rarely incorporated novel biomarker assessments into their approaches. In fact, the subarea of sociology related to health issues generally was identified as medical sociology and was primarily oriented toward understanding the ways in which healthcare organization and patient behaviors were dynamically intertwined. All this has changed over the last fifteen years. The greatest strides in this area have come about because vigorous sociocentric approaches—approaches in which entire networks have been mapped—have been incorporated into large-scale studies with strong health assessments. Furthermore, demographers during this period have become much more centrally involved in understanding socioeconomic conditions, family dynamics, and mortality. Both epidemiologists and sociologists have incorporated each other's strongest approaches with growing success. Notable among the studies that have integrated social network assessments, health, and biomarkers are the Framingham Study, the National Longitudinal Study of Adolescent Health (Add Health), a number of HIV-related studies, and an increasing presence of sociocentric methods in the prevention programs and in aging studies in Europe (the Survey of Health, Ageing and Retirement in Europe [SHARE] and the English Longitudinal Study of Ageing [ELSA]). These new studies have used both egocentric models in which only direct links to individual participants are identified as well as sociocentric models of entire networks where the spread of disease and behaviors can be studied. During this same period, a large body of work produced by social psychologists on social isolation and loneliness has also flourished and has changed the landscape, reinforcing the importance of perceptions in mediating impacts on health and well-being. This chapter is devoted to incorporating findings from these studies and methods into mainstream discussions in social epidemiology.

Our ability to identify the causal impact of social relations has also increased substantially in the last 10–15 years. In large part, this is the result of much more nuanced work on issues of causal inference building on recent methodological advances in observational studies and the analysis of several large-scale randomized experiments (1–3). The results of many of the randomized trials in which morbidity or mortality were outcomes have yielded null or very weak results—calling into question the impact of social network interventions. These findings challenge the notion of a causal impact *or* our ability to actually alter networks and support during important etiologic periods. New work incorporating lifecourse approaches has also begun to flourish, and we see that social relationships, not surprisingly, are often formed in childhood or early adulthood, and relational skills are built even earlier. Thus, it is critical to incorporate lifecourse approaches into network epidemiology. Although this issue has been raised frequently enough, there is still less work in this area in contrast to the fuller incorporation of stronger network assessments. The specifics of the network/support intervention randomized controlled trials (RCTs) and contested areas of causal inference are discussed in Chapter 11 on psychosocial interventions.

The evidence linking the influence of social relationships—defined broadly as the degree to which individuals are interconnected and embedded in communities—on health and longevity is now enormous. Furthermore, the disciplinary divisions among social epidemiology, sociology, and social psychology are increasingly blurred as all disciplines incorporate both observational and experimental designs into their work and add physiological and clinical assessments to large and small studies alike. A rich combination of observations and theoretical literature on social integration, attachment, and social networks led us originally to test these ideas empirically. Humans are social animals; the need for intimacy, nurturance, and connection is built into our being. Now, over 35 years after John Cassel (4), Sidney Cobb (5), and other seminal thinkers in social epidemiology suggested that this was a critical area of investigation, and 20 years after the earliest studies in Alameda County, California; Tecumseh, Michigan; and Durham County, North Carolina, revealed the influence of social relationships on mortality (6–8), it is time to take stock of the vast literature on this topic. A recent meta-analytic review identified 148 studies on the topic of social ties and mortality (9). Our aim is to revisit some of the seminal theories that have guided empirical work, and to revise and reformulate some of those ideas, especially in light of the more recent sociocentric approaches and results of RCTs, and to point the way toward productive lines of inquiry for the future.

When investigators write about the impact of social relationships or, more specifically, of social networks on health, many terms are used loosely and interchangeably, including social networks, social support, social isolation, and social integration. A major aim of this chapter is to define and clarify these terms. We discuss *(1)* theoretical orientations from diverse disciplines that are fundamental to advancing research in this area; *(2)* an overarching model that integrates multilevel systems or structures; *(3)* a set of definitions accompanied by major assessment tools; *(4)* some of the strongest findings linking social networks or support to morbidity, mortality, or functioning; and finally, *(5)* a series of recommendations for future work. Since there are now numerous books and literature reviews on networks, support, and health (10–23), our aim is not to be all-inclusive but rather to highlight work that has substantially advanced our thinking in this area and to give the reader a sense of the range and depth of this literature, now a body several decades in the making.

THEORETICAL ORIENTATIONS

Several theories form the bedrock for the empirical investigation of social relationships and their influences on health. The earliest theories came from sociologists such as Émile Durkheim as well as from psychoanalysts such as John Bowlby, who first formulated attachment theory. A major wave of conceptual development came from anthropologists, including Elizabeth Bott, John Barnes, and Clyde Mitchell, and quantitative sociologists such as Claude Fischer, Edward Laumann, Barry Wellman, and Peter Marsden, who, along with others, developed social network analysis. This eclectic mix of theoretical approaches, coupled with the work on stress early on by Cannon and Selye and later by McEwen, Cohen, and Cacioppo (24–32), addresses the protective roles of social resources and support within the context of research on stress. Combined with this work, the contributions of social epidemiologists John Cassel and Sidney Cobb form the foundation of research on social ties and health.

SOCIAL NETWORK ANALYSIS: A NEW WAY OF LOOKING AT SOCIAL STRUCTURE AND COMMUNITY

During the mid-1950s, a number of British anthropologists found it increasingly difficult to understand the behavior of either individuals or groups on the basis of traditional categories such as kin groups, tribes, and villages. Barnes (33) and Bott (34) developed the concept of "social networks" to analyze ties that cut across traditional kinship, residential, and class groups to explain behaviors they observed such as finding jobs, political activity, and social roles. The development of social network models provided a way to view the structural properties of relationships among people with no constraints or expectations that these relationships occurred only among bounded groups defined a priori.

As this work and the work of other European post–World War II sociologists became known in the United States, American sociologists extended the concept of social network analysis, incorporating into it their more quantitative orientation. Wellman (35), in several historical reviews of the development of social network analysis, has described "the network" of network analysis. A strong center started at Harvard under Harrison White and Charles Tilly and extended to their graduate students: Edward Laumann (36) went to the University of Chicago, Barry Wellman (37) to the University of Toronto, and Mark Granovetter (38) and Claude Fischer (39, 40) to the University of California, Berkeley. These sociologists developed what has come to be known as an egocentric network approach to social network analysis, in which the structure and function of networks are assessed from the perspective of an individual. Network analysis "focuses on the characteristic patterns of ties between actors in a social system rather than on characteristics of the individual actors themselves. Analysts search for the structure of ties underlying what often appears to be incoherent surface appearances and use their descriptions to study how these social structures constrain network members' behavior" (41). Network analysis addresses the structure and composition of the network and the contents or specific resources that flow through those networks. Social network analysis includes analyses

of both egocentric networks with an individual at the center and entire sets of networks at the level of communities or workplaces. The analysis of entire networks employs sociocentric approaches, the study of entire bounded communities where network relationships of entire schools, towns, or workplaces are identified.

The strength of social network theory rests on the testable assumption that the social structure of the network itself is largely responsible for determining individual behavior and attitudes by shaping the flow of resources or information that determine access to opportunities and constraints on behavior. Network theorists share many of the central assumptions of Durkheim and the structural functionalists. The central similarity is the view that the structure of social institutions shapes the resources available to the individual and hence that person's behavioral and emotional responses. Another contribution of network theory is the observation, initially made by Barnes and Bott that the structure of networks may not always conform to preconceived notions of what constitutes "community" defined on the basis of geographic or kinship criteria. Thus, Wellman argues that the essence of community is its social structure, not its spatial structure (42). By assessing actual ties between network members, one can empirically test whether community exists and whether that community is defined on the basis of neighborhood, kinships, friendship, institutional affiliation, or other characteristics. This emphasis is shared by Durkheim (43), who describes a shift from mechanical solidarity (based on kinship ties) to organic solidarity (based on rational exchange-based ties) as the basis of social organization.

SOCIAL INTEGRATION, ALIENATION, AND ANOMIE: DURKHEIM'S CONTRIBUTION

> Suicide varies inversely with degree of integration of the social groups of which the individual forms a part.
>
> —Émile Durkheim (44)

Émile Durkheim, a French sociologist working late in the nineteenth century, was one of the founding fathers of sociology. Durkheim's contribution to the study of the relationship between society and health is immeasurable. Perhaps most important is the contribution he made to the understanding of how social integration and social cohesion influence mortality. Durkheim's primary aim was to explain individual pathology as a function of social dynamics. In light of emerging attention to "upstream" determinants of health in the mid-1990s (45), Durkheim was indeed ahead of his time.

While a professor at the University of Bordeaux, Durkheim wrote three of his four most important books: *The Division of Labor in Society* (43), *The Rules of Sociological Method* (46), and *Suicide* (44). *Suicide* lays the framework for understanding the role of social integration in health. Building on *The Rules of Sociological Method*, Durkheim challenges himself to understand how the patterning of one of the most psychological, intimate, and, on the surface, individual acts rests on the patterning of "social facts." As noted by Bierstedt (47), it is as if Durkheim chooses for himself the hardest of challenges to prove the power of social phenomena to influence what seem to be individual acts.

In *Suicide*, Durkheim shows how "social facts" can be used to explain changing patterns of aggregate tendency toward suicide. He argues that individuals are bonded to society by two forms of integration: attachment and regulation. Attachment is the extent to which an individual maintains ties with members of society. Regulation involves the extent to which an individual is held in the fabric of society by its values, beliefs, and norms (48). Because Durkheim's logic and language are so elegant, in the following paragraphs we try to give the reader the flavor of his thinking as it relates to social integration and suicide.

Durkheim starts his work with the observation that countries and other geographic units and social groups have very stable rates of suicide year after year:

> Thus, individuals making up a society change from year to year, yet the number of suicides itself does not change . . . the population of Paris renews itself very rapidly, yet the share of Paris in the total number of French suicides remains practically the same . . . the rate of military suicides varies only very slowly in a given nation. . . . Likewise, regardless of the diversity of individual temperaments, the relation between aptitude for suicide of married persons and that of widowers and widows is identically the same in widely differing social groups. The causes which thus fix the contingent of voluntary deaths for a given society or one part of it must then be independent of individuals, since they retain the same intensity no matter what particular persons they operate on. (44)

Durkheim's contribution to our understanding of how social structure—and particularly levels of integration based on religious, family, and occupational organization—affects suicide is unparalleled. He paved the way for much of the work in this area through the development and testing of basic sociological theories that have largely survived the test of time. He viewed suicide not as an "isolated tragedy" in the life of an individual but as a reflection of conditions of society as a whole (49).

ATTACHMENT THEORY ACROSS THE LIFECOURSE: BOWLBY'S CONTRIBUTION

> All of us from the cradle to the grave are happiest when life is organized as a series of excursions, long or short, from the secure base provided by our attachment figures.
>
> —John Bowlby (50)

John Bowlby has been described as one of the most important psychiatrists in the twentieth century (51). He qualified as a psychoanalyst in 1937 and soon thereafter was proposing theories to the British Psychoanalytic Society suggesting that the environment, especially in early childhood, plays a critical role in the genesis of neurosis. Early in his career, he believed that the separation of infants from their mothers was unhealthy. He saw loss and separation as key issues for psychotherapy. Bowlby proposed that there is a universal human need to form close affectional bonds (52). Between 1964 and 1979, Bowlby wrote a major trilogy, *Attachment* (53), *Separation* (54), and

Loss (55), in which he laid out his theory of attachment and how it relates to both childhood and adult development.

Attachment theory contends that the attached figure—most often, but not necessarily, the mother—creates a secure base from which an infant or toddler can venture forth and explore. Bowlby argued with many psychoanalysts that attachment is a "primary motivational system" (i.e., not secondary to feeding or warmth) (53). "Secure attachment," he wrote, "provides an external ring of psychological protection which maintains the child's metabolism in a stable state, similar to the internal homeostasis mechanisms of blood pressure and temperature control" (53). These intimate bonds created in childhood form a secure base for solid attachment in adulthood and provide prototypes for later social relations (52). Secure attachment, as opposed to avoidant, ambivalent, or disorganized attachment, allows the maintenance of affectional bonds and security in a larger system. We now have increasing evidence of the importance of such early attachment for emotional regulation and adult health (56–64).

In adulthood, Bowlby saw marriage as the adult equivalent of attachment between infant and mother during childhood. If secure, marriage would provide a solid base from which to work and explore the world enmeshed in a "protective shell in times of need" (65).

The strength of Bowlby's theory lies in its articulation of an individual's need for secure attachment for its own sake, for the love and reliability it provides, and for its own "safe haven." Primary attachment promotes a sense of security and self-esteem that ultimately provides the basis on which the individual will form lasting and loving relationships in adult life. The ideas of attachment and loneliness are currently expressed in much of the work on loneliness by Cacioppo (12–14, 18, 66, 67). The psychosocial environment in infancy and childhood paves the way for successful development that continues through adulthood. For Bowlby, the capacity for intimacy in adult life is not a given; it is instead the result of complex dynamic forces involving attachment, loss, and reattachment. Throughout this volume, we have seen the growing importance of bringing such a lifecourse and dynamic perspective to understanding social determinants of disease.

WEAVING THE THREADS TOGETHER

How do these theories from very different perspectives come together to help us develop a conceptual framework with which to examine the ways social relationships influence health? How can we hope to integrate a set of ideas proposed by sociologists, anthropologists, and psychiatrists writing over the last century, none of whom was interested primarily in the broad array of health outcomes falling under the purview of the epidemiologist? To begin with, we draw from these theorists the greatest contributions to the development of a comprehensive framework of use to social epidemiology. For instance, a singular contribution of Durkheim's was his anchoring of an individual's risk of death in the social experience of the group. His steadfast orientation toward population patterns of mortality permitted him to identify social integration as a critical contributor to the social patterning of suicide. Without denying that the characteristics of the individual or proximate and precipitating factors could influence *who* among many in a particular group might commit suicide, his constant orientation to population patterns allowed him to uncover collective, societal characteristics related to suicide. Bowlby's view of attachment as a "primary

motivational system" is critical because attachment provides love, security, and other nonmaterial resources—not only food, warmth, and material resources. The work of Charles Nelson on Romanian orphans documents this idea in a compelling way (64). This theory is also central to our thinking of the way in which social relationships may be health promoting. Bowlby tried to identify critical periods in development when bonds of attachment are made. This lifecourse perspective has flourished in social epidemiology over the last decade (68, 69). Finally, much of our framework builds directly on the work of social network theorists. Critical contributions center on the network approach itself, in which the structure and function of ties are assessed without the assuming they are defined by specific kinds of "bounded" affiliations such as kin, neighborhood, and work. This orientation permitted social network analysts from Bott (34) to Wellman (42) to identify the social structure underlying behaviors when a traditional focus on either family or neighborhood was incapable of explaining behavioral patterns.

Two other strengths of network theories deserve mention. First, the flexibility of social network models in spanning assessment of intimate as well as extended ties permits a deep understanding and appreciation of the critical roles many kinds of relationships play in everyday life. Second, network theories virtually force researchers to identify characteristics of the network (at the social level) rather than characteristics of the individual as explanatory variables. Thus we see structural network characteristics explaining support, access to jobs (38, 70), social influence (19–21, 71), health behaviors (72–74), and disease transmission (75–79). By integrating these diverse theories and weaving them together, we derive powerful theories and models. The work of Christakis, Bearman, Moody, Morris, and Valente (19–21, 73, 74, 76, 80–96) in the United States and Kohler, Watkins, and colleagues (74, 78, 79, 97–102) in Africa illustrates the critical insights that are gained from this more formal network approach. We use them to build a more comprehensive framework, which we use to examine how social relations and networks influence a broad array of health outcomes.

A CONCEPTUAL MODEL LINKING SOCIAL NETWORKS TO HEALTH

AN OVERVIEW

Beginning with seminal work in epidemiology by Cassel (4) and Cobb (5), who first suggested a link between social resources, support, and disease risk, epidemiologists began to investigate the role of social relationships on health. Throughout the 1970s and 1980s a series of studies appeared that consistently showed that the lack of social ties or social networks predicted mortality from almost every cause of death (see reviews 10, 24, 103). These studies most often captured numbers of close friends and relatives, marital status, and affiliation or membership in religious and voluntary associations. These measures were conceptualized in any number of ways as assessments of social networks or ties, social connectedness, integration, activity, or, conversely, social isolation. Whatever they were named, they uniformly defined integration as involvement with ties spanning the range from intimate to extended ties. Most studies included measures of both "strong"

and "weak" ties. As defined by Mark Granovetter (38), weak ties involve contacts with extended nonintimate ties, which he found to be central to occupational mobility.

Although the power of these measures to predict health outcomes is indisputable, the interpretation of what the measures actually assess has been open to much debate. Hall and Wellman (41) have appropriately commented that much of the early work in social epidemiology had used the term "social networks" metaphorically, since rarely have investigators conformed to more standard assessments used in network analysis. For instance, the existence of "weak ties" is not assessed directly, but inferred from membership in voluntary and religious organizations. This criticism was taken to heart, and another generation of network measures was developed that incorporated more dimensions of networks and functions (24, 103–105).

A second wave of research developed in reaction to this early work and as an outgrowth of work in psychology that changed the orientation of the field in several ways. Major contributors to this second wave include Antonucci (105–107), Kahn (108), Lin (109–113), House (114–116), Cohen (117–121), Rook (122–124), and Barbara and Irwin Sarason (125–126). These social scientists focused on the provision of social support rather than on the elaboration of the structural aspects of social networks. Especially important among these contributions was Kahn and Antonucci's formulation of the convoy model, in which the individual is seen in a lifecourse perspective as traveling through life surrounded by members of his/her cohort who share experiences and life histories and who provide support to one another reciprocally over time (107, 128).

Our understanding of the richness and complexity of social support has been advanced immeasurably by Lin's resource theory, definitions of support developed by House and Kahn, and Sarason's call for more theory-based work. They have helped us understand how support is linked to mental health. But these investigators most often share an assumption—that the key function of social networks is provision of social support. Social support is one of the main ways social networks influence physical and mental health status. We now understand, however, that support is not the only critical pathway. Moreover, the exclusive study of more proximal pathways detracts from the need to focus on the social context and structural underpinnings in which social support is provided. In order to have a comprehensive framework within which to explain these phenomena, we must move "upstream" and focus on network structure. Only then can we fully consider the multiple pathways by which social networks might profoundly influence health outcomes. It is also critical to maintain a view of social networks as lodged within those larger social and cultural contexts that shape the structure of networks.

In recent years, a third generation of network and health studies has appeared. These studies have the great advantage of drawing on formal network analysis using both egocentric and sociocentric models and relying on mathematical models to describe network structure, transmission of disease, behavior, and attitudes. They move from studies of social support to social networks. The study of networks per se is broader than the study of support because conceptually networks have "emergent properties not explained by the constituent parts" (23, 129). These new studies cover network ties and health behaviors (19, 22, 73, 81–87, 90, 92, 95, 130) in the United States. Other investigators have developed studies of HIV/AIDS transmission related to sexual networks and have made seminal contributions to network epidemiology (76, 78, 79, 131).

In Figure 7.1, we present a conceptual model of how social networks impact health. We envision a cascading causal process beginning with the macrosocial to psychobiological processes that are dynamically linked together to form the processes by which social integration affects health.

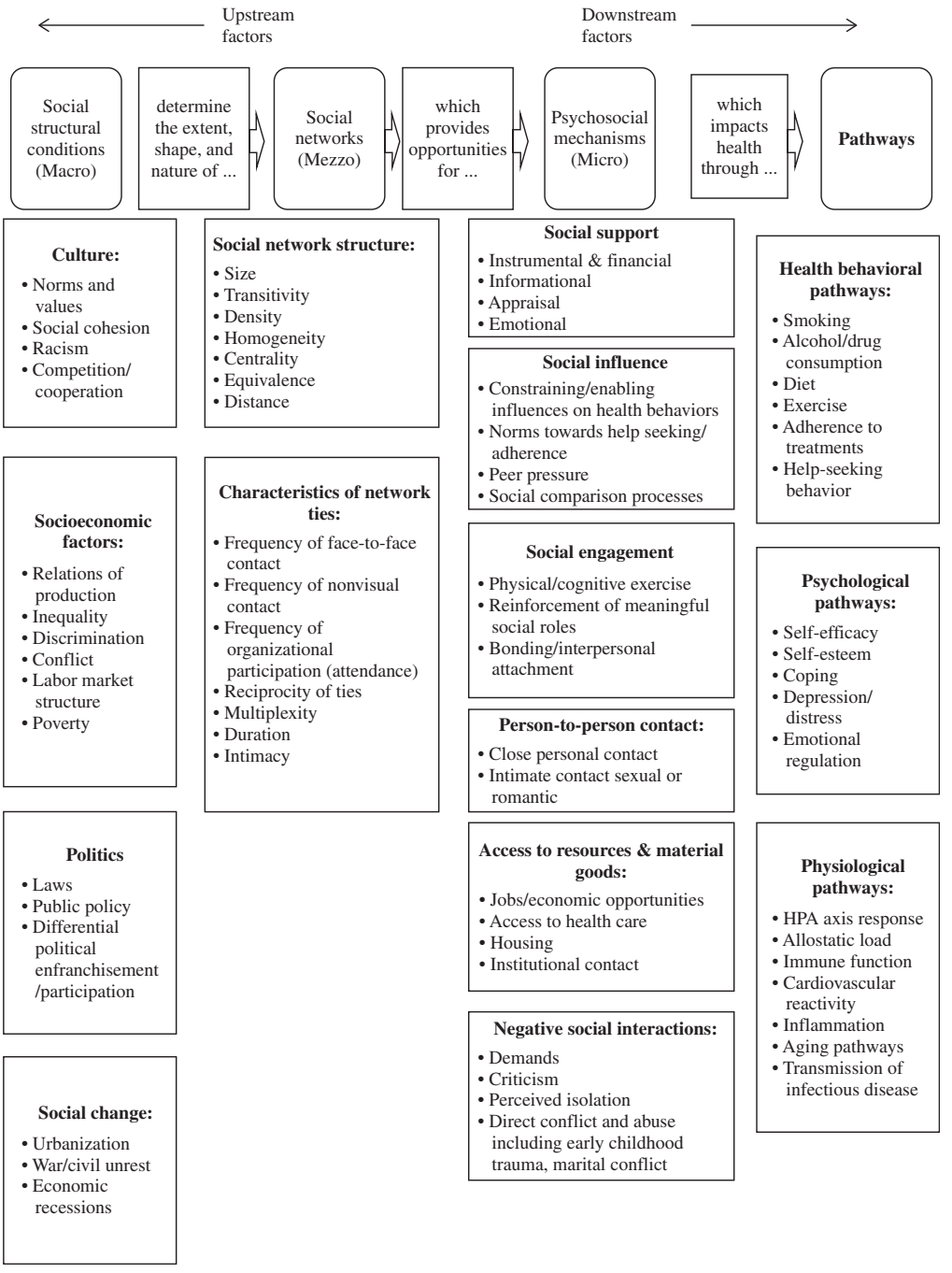


FIGURE 7.1: Conceptual models of how social networks impact health.

As suggested above, we start by embedding social networks in a larger social and cultural context in which upstream forces are seen to condition network structure. Serious consideration of the larger macrosocial context in which networks form and are sustained has been lacking in all but a small number of studies and is almost completely absent in studies of social network influences on health.

We then move downstream to understand the influences network structure and function have on social and interpersonal behavior. We argue that networks operate at the behavioral level through five primary pathways: *(1)* provision of social support; *(2)* social influence; *(3)* social engagement and attachment; *(4)* access to resources and material goods; and *(5)* negative social interactions including conflict and abuse. These micropsychosocial and behavioral processes, we argue, then influence even more proximate pathways to health status. These include direct physiological stress responses; health-damaging behaviors such as tobacco consumption or high-risk sexual activity; health-promoting behavior such as appropriate health service utilization, medical adherence, and exercise; and finally, exposure to infectious disease agents such as HIV, other sexually transmitted diseases (STDs), or tuberculosis. A fuller description of the biological pathways by which social relationships might influence health across the lifecourse is discussed in Chapter 14 on biological embedding. Here we provide a more abbreviated review.

By embedding social networks in this larger chain of causation, we integrate more fundamentally "upstream" macrosocial forces related to the political economy with social networks as mediating structures between the largest- and smallest-scale social forms. Thus, we can examine how labor markets, economic pressures, and organizational relations influence the structure of networks (132–135). We can examine specifically how culture, rapid social change, industrialization, and urbanization affect the structure of networks. Perhaps the most critical findings to date in this area relevant to social epidemiology are whether "community" is dead or dying in postindustrial American society. In fact, this question has been central to many social network analysts (42, 136, 137).

DOWNSTREAM SOCIAL AND BEHAVIORAL PATHWAYS

SOCIAL SUPPORT

Moving downstream, we now come to a discussion of the mediating pathways by which networks might influence health status. Most obviously, the structure of network ties influences health via the provision of many kinds of support. This framework immediately acknowledges that *not all* ties are supportive and that there is variation in the type, frequency, intensity, and extent of support provided. For example, some ties provide several types of support while other ties are specialized and provide only one type.

Social support is typically divided into subtypes that include emotional, instrumental, appraisal, and informational support (24, 114, 118, 121, 138, 139). Emotional support is related to the amount of "love and caring, sympathy and understanding and/or esteem or value available from others" (140). Emotional support is often provided by a confidant or intimate relationship, although less intimate ties can provide such support under circumscribed conditions.

Instrumental support refers to help, aid, or assistance with tangible needs such as getting groceries, getting to appointments, phoning, cooking, cleaning, or paying bills. House (114) refers to instrumental support as aid in kind, money, or labor. Appraisal support, often defined as the third type of support, relates to help in decision-making, giving appropriate feedback, or help deciding which course of action to take. Informational support is related to the provision of advice or information in the service of particular needs. Emotional, appraisal, and informational support are often difficult to disaggregate and have various other definitions (e.g., self-esteem support).

We share the view of Kahn and Antonucci (128), who view social support as transactional in nature, potentially involving both giving and receiving. Further, the process of giving and receiving support resources occurs within a normative framework of exchange in which behavior is guided by norms of interdependence, solidarity, and reciprocity (see 141). Support exchanges also take place within a lifecourse context and not simply in response to day-to-day contingencies. This helps explain patterns of continued support exchange in late life among persons who are disabled and unable to reciprocate. Moreover, support exchanges take place within the context of social network ties, which are long-standing and based on shared histories and not as isolated or atomized phenomenon. Measures of support frequently fail to assess such aspects of reciprocity and instead focus more attention on received support.

Apart from type of support, it is important to differentiate cognitive from behavioral aspects of support. That a person perceives support to be available on need may or may not correspond with the actual provision of that support in circumstances in which such a request is made. Both the cognition of one's sense of the availability and adequacy of potential support and the extent to which support is actually received appear to be different and equally important. Support that is received is an actual exchange related to a behavior. It is sometimes called enacted or experienced support (142). A brisk debate persists over which is more important in what situations—behavioral or cognitive—in either case, it is clear that they tap different aspects of support and are only modestly correlated in most studies (142).

Unlike emotional support, instrumental, appraisal, and informational support may influence health because these types of support improve access to resources and material goods. Classic examples would be Granovetter's study of the strength of "weak ties," in which ties that are personally less intimate but that bridge across networks provide for better access to jobs (38). Support conceived of in these ways provides economic opportunity and access to healthcare, and creates institutional liaisons.

SOCIAL INFLUENCE

Networks may influence health via several other pathways. One pathway that is often ignored is based on *social influence.* Marsden asserts that the "proximity of two actors in social networks is associated with the occurrence of interpersonal influence between the actors" (71). As the term is used, influence need not be associated with face-to-face contact, nor does it require deliberate or conscious attempts to modify behavior (71). Marsden refers to work by Erickson (143) suggesting that under conditions of ambiguity "people obtain normative guidance by comparing their attitudes with those of a reference group of similar others. Attitudes are confirmed and reinforced when they are shared with the comparison group but altered when they are discrepant" (71).

Shared norms around health behaviors (e.g., alcohol and cigarette consumption, healthcare utilization, treatment adherence, and dietary patterns) might be powerful sources of social influence with direct consequences for the behaviors of network members. These processes of mutual influence might occur quite apart from the provision of social support taking place within the network concurrently. The classic study by Christakis and Fowler document the network influences on tobacco consumption (73) and obesity (87). A number of studies from the National Longitudinal Study of Adolescent Health (Add Health) document peer influences (22, 81, 83, 84, 92). Perceptions of risk and attitudes toward HIV/AIDS are other recent examples of the ways in which social influence shapes health (79). The social influence that extends from the network's values and norms constitutes an important and underappreciated pathway through which networks impact health.

SOCIAL ENGAGEMENT

A third pathway by which networks may influence health status is by promoting social participation and social engagement. Participation and engagement result from the enactment of potential ties in real-life activity. Getting together with friends, attending social functions, participating in occupational or social roles, group recreation, religious attendance—these are all instances of social engagement. Thus, through opportunities for engagement, social networks define and reinforce meaningful social roles including parental, familial, occupational, and community roles, which in turn provide a sense of value, belonging, and attachment. Those roles that provide each individual with a coherent and consistent sense of identity are only possible because of the network context, which provides the theater in which role performance takes place.

In addition, network participation provides opportunities for companionship and sociability. We, as well as others (122), argue that these behaviors and attitudes are not the result of the provision of support per se but are the consequence of participation in a meaningful social context in and of itself. We hypothesize that part of the reason measures of social integration or "connectedness" have been such powerful predictors of mortality for long periods of follow-up is that these ties give meaning to an individual's life by enabling him or her to participate in it fully, to be obligated (in fact, often to be the provider of support), and to feel attached to one's community. Despite the tendency of some researchers to classify "belonging" as another feature of support, this pathway is distinct from the level of support that is either received or even perceived, standing apart from cognitive and behavioral aspects of support. Such a pathway relates closely to the way in which social networks contribute to social cohesion. Through contact with friends and family and participation in voluntary activities, life acquires a sense of coherence, meaningfulness, and interdependence.

An impressive amount of research now links social engagement and participation to cognitive function in old age. The mechanism linking social engagement may also include the direct stimulation of cognitive functions such as executive function (144–146). This is discussed in further detail in a separate section on cognitive function in this chapter. Thus, social engagement may activate physiologic systems that operate directly to enhance health as well as indirectly by contributing to a sense of coherence and identity that allows for a high level of well-being.

PERSON-TO-PERSON CONTACT

Networks also influence disease by restricting or promoting exposure to infectious disease agents. In this regard the methodological links between epidemiology and networks are striking. Furthermore, as we distinguish the pathways leading from network structure to health, we find that the networks can be both health-promoting and health-damaging if they serve as vectors for the spread of infectious disease and simultaneously provide emotional support. Efforts to link mathematical modeling by applying network approaches to epidemiology have advanced substantially over the last 10 years (81, 82, 96, 98, 131, 147–154). In an insightful paper, Morris (148) discusses how epidemiologists developed models of disease transmission by initially recognizing the biological characteristics of the disease agent. By the turn of the century, epidemiologists had recognized that the population dynamics of an epidemic are proportional to *(1)* the probability that one member of the contact is susceptible; *(2)* the probability that the other is infected; and *(3)* the number of effective contacts made between individuals per unit time (149). The contribution of social network analysis to the modeling of disease transmission is the understanding that in many, if not most, cases, disease transmission is not spread randomly throughout a population. Social network analysis is well suited to the development of models in which exposure between individuals (for instance, sexual networks) is not random but rather is based on geographic location, demographic characteristics (age, race, gender), or other important characteristics of the individual (socioeconomic position, occupation, sexual orientation) (81–83, 150). Furthermore, because social network analysis focuses on characteristics of the network rather than on characteristics of the individual, it is ideally suited to the study of diffusion of transmissible diseases through populations via bridging ties between networks or to uncovering characteristics of ego-centered networks that promote the spread of disease.

Perhaps the most successful example to date of the application of network analysis to the spread of infectious disease is work done on HIV transmission. Whether spread through sexual contact or intravenous drug use, HIV transmission results from selective rather than random mixing. HIV/AIDS in the United States as well as in Sub-Saharan Africa today is perhaps best understood from a network perspective. Understanding the dynamics of disease spread predominantly by person-to-person contact requires an appreciation for the complex dynamics between individuals and their social networks.

ACCESS TO MATERIAL RESOURCES

Surprisingly little research has sought to examine differential access to material goods, resources, and services as a mechanism through which social networks might operate. This, in our view, is unfortunate given the work of sociologists showing that social networks operate by regulating an individual's access to life opportunities by virtue of the extent to which networks overlap with other networks. Perhaps the most important among these studies is Granovetter's (38) classic study of the power of "weak ties" that, on the one hand, lack intimacy, but on the other hand facilitate the diffusion of influence and information and provide opportunities for mobility.

We speculate that participation in networks on the basis of shared work experiences (e.g., trade unions, professional organizations), health experiences (support groups for recovery from cancer,

stroke, heart disease), or religious affiliation provides access to resources and services that have a direct bearing on health outcomes. Quite apart from the support provided by these ties—even the instrumental support provided—membership in these groups may provide access to job opportunities, high-quality health care, and housing. While this pathway is closely allied with instrumental, appraisal, and financial support, we believe that further empirical work and increased understanding may show that it constitutes a linkage between networks and health not defined primarily by support.

NEGATIVE INTERACTIONS: CONFLICT AND STRAIN

The downside of social relationships includes the demands, criticism, perceived isolation, and direct conflict and abuse that relationships can produce (146, 155–158). Negative social interactions are well known to influence a number of physiological stress pathways. A literature on early childhood trauma stemming from verbal and physical abuse, neglect, and lack of love and affection during childhood has shown long-term adult sequelae for both physical and mental health (61, 159–162). In experimental laboratory studies, negative interactions may produce immediate physiological reactions. Among adults, marital quality, conflict, and demands influence inflammatory processes (163, 164), as well as a host of cardiovascular risk factors and changes in cortisol. These stressful experiences are, like other major stressors, tied to mortality and morbidity risks. Antonucci and others find that negative interactions occur most often among network members who are also intimate ties (165–167).

We have identified five mechanisms by which the structure of social networks might influence disease patterns. Social support is the mechanism most commonly invoked, but social networks also influence health through additional mechanisms, including forces of social influence, levels of social engagement and participation, regulation of contact with infectious disease, access to material goods and resources, and negative interactions. These mechanisms are not mutually exclusive. In fact, it is most likely that in many cases they operate simultaneously. The researcher starting an investigation in this area needs to develop clear hypotheses a priori about aspects of network structure and the mechanism(s) through which it may influence health to maximize opportunities to understand the way in which social structures are linked to health.

BIOLOGICAL AND PSYCHOLOGICAL PATHWAYS PROXIMATE TO HEALTH STATUS

Social networks operate through the previously described series of five mechanisms in shaping the health of individuals. In turn, these behavioral mechanisms affect other downstream factors via biological and psychological pathways most proximate to the health outcome. Moving across our diagram (see Figure 7.1), we now turn our attention to these pathways. Three distinct pathways will be outlined, although, again, the reader is alerted to the distinct possibility, in fact, the likelihood, that multiple pathways are involved simultaneously.

First, social networks via social influence or supportive functions influence health-promoting or health-damaging behaviors such as tobacco and alcohol consumption, physical activity, dietary patterns, sexual practices, and illicit drug use. Second, social networks via any number of pathways influence cognitive and emotional states such as self-esteem, social competence, self-efficacy, depression, and affect. Third, networks may have direct effects on health outcomes by influencing a series of physiologic pathways largely related to stress responses. (See Chapter 14 on biological embedding for a fuller discussion of these pathways.) More recently, biological pathways related to restorative functions and resilience have been documented complementing the negative, stressful pathways. The reader is referred to two excellent recent reviews on the physiologic and behavioral processes linked to social networks and support (168, 169) and to several more linking social isolation to physiologic outcomes (12, 17, 18, 66, 67, 170–172).

HEALTH BEHAVIORS

Social networks influence patterns of risk-related or health-promoting behaviors including tobacco, alcohol, and other substance use, patterns of physical activity and diet, and behaviors related to sexual activity. Networks provide opportunities for sharing behaviors, norms around such behaviors, and support for behavioral decisions. Peers are profoundly important in the lives of adolescents when initiation of many behaviors occurs and again in adulthood when changes related to cessation (alcohol or tobacco use) or health promotion present themselves. Work on the Add Health Study (22, 81, 83, 84, 92, 173–175) reveals the enduring importance of adolescent networks. In higher risk groups, synergetic effects of networks on multiple high-risk behaviors are also reported (176). In addition, cohesive, supportive networks may themselves blunt stressful experiences and enable people to resist risky behavior and maintain healthier choices. Over the last 10 to 15 years, the literature in this area has mushroomed. Here, we discuss a handful of landmark stories that have altered the ways in which we understand how social networks influence health behaviors.

The social network landscape changed with the analysis of the Framingham Heart Study by Christakis and Fowler (19, 22, 72, 73, 87). In a series of papers on smoking, alcohol consumption, obesity, and other health risks, they show the network dynamics in a sociocentric study of participants in the Framingham Heart Study from 1971 to 2003. In the 2008 paper on smoking, Christakis and Fowler report that the average risk of smoking is 61% higher if the contact is closely tied to a smoker by one degree of separation (73). This percentage decreased to 29% for contacts with two degrees of separation from a smoker and further decreased to 11% for those with three degrees of separation. With four degrees of separation there was no excess risk. Furthermore, whole clusters of participants became nonsmokers within the same period, suggesting that smoking cessation was in some way a collective phenomenon. Finally, coworkers in small firms were willing to stop smoking together, as were spouses. Between 1970 and 2000, smokers became increasingly marginalized in their networks. These findings have important implications for prevention programs in which network structure can be incorporated into a number of behavior change programs where the diffusion of innovation is central (94, 95, 177).

PSYCHOLOGICAL MECHANISMS

Social relations ranging from close and intimate family ties from early childhood to adulthood as well as the wider set of relations in extended social networks shape our affective and cognitive states. In the first edition of this book, much of the attention was focused on self-efficacy. More recently, attention has centered on early family and parental experiences as they relate to the capacity for emotional regulation in early childhood and lead to a number of emotional states and cognitive strategies in later adulthood (see Chapter 9 on affective states for a fuller discussion). Chapter 9 discusses emotions related to compassion and gratitude, which surely stem in part from close bonds and may also impact health. And negative emotions such as depression have long been related to social support in bidirectional ways (178–181). Here we again focus on the potential for self-efficacy to mediate the impact of social networks on health outcomes. The interplay between emotions and social affiliation is clearly dynamic with psychologists finding support for the theory that positive affective states foster social support and capital (182, 183). Here we continue to focus on the potential for networks to promote self-efficacy. Self-efficacy, defined as the degree of confidence people have in their ability to perform specific behaviors, has been shown to be associated with a variety of health and functional outcomes (184–188). A considerable body of evidence undergirds the assertion that self-efficacy is one of the psychosocial pathways through which social support operates. For example, in a study of postpartum depression, the protective effect of social support was observed to occur primarily through its mediation of maternal feelings of self-efficacy (189). Other studies have observed the indirect influence of social support through enhanced self-efficacy in coping with abortion (190), smoking cessation (191), and depression (192). The association between social networks and health-promoting behavior such as exercise has also been shown to be mediated through self-efficacy (193).

Evidence suggests that ongoing network participation is essential for the maintenance of self-efficacy beliefs in late life. A study by McAvay et al. (194) found that lower levels of social network contact were predictive of decline in the health and safety domains of self-efficacy; the absence of instrumental support was also associated with decline in the productivity, health, and transportation domains. There is some evidence that the impact of self-efficacy and social support are reciprocal—meaning that while social support may bolster self-efficacy, it may also be the case that self-efficacy is independently associated with higher levels of social support (195). The complexity of these reciprocal dynamics has yet to be fully examined.

In addition to self-efficacy, social integration appears to operate through additional psychosocial pathways. For example, some evidence suggests that social support promotes functional and adaptive coping styles (195, 196). An influential study by Dunkel-Schetter et al. (197) has shown, however, that these relationships are likely to be reciprocal. Their evidence suggests that in stressful situations, different coping styles elicit different responses from the social environment. Indeed, the tendency to ask for and make use of social support itself is one of many possible coping styles, and has numerous psychological antecedents and correlates (198). In a review of patterns of attachment, Fonagy (52) presents evidence that attachment relationships contribute to self-esteem and the perception that the individual is in control of his or her own destiny.

Social support may additionally operate through its influence on emotion, mood, and perceived well-being. Numerous studies have shown that social support is associated with symptoms of depression (109, 199–209). This evidence is particularly important in light of the fact that

social support—especially perceived emotional support—has been shown to buffer the deleterious influences of stressful life events on the risk of depression and depressive symptoms (112, 210, 211). The evidence appears to be strong that those who are socially isolated are at increased risk of depression, especially in late life (212). The relationship in some cases is reciprocal, with support influencing depressive symptoms and vice versa (206). In studies of psychological health, one consistent finding is that the perceived adequacy of social support, more so than the availability of support, appears to be most important (213).

PHYSIOLOGICAL PATHWAYS

An examination of the pathways linking social networks to health outcomes yields a rich and complex latticework of interlinking mechanisms—biological, psychological, and physiological—that cascades from the macro to the micro, from upstream to downstream, to generate potentially powerful influences on health and well-being across the lifecourse. One of the robust findings in the literature on networks and health is the broad impact that network integration has on all-cause mortality. This may be related to the numerous pathways, which more proximately affect disease onset or progression, but it is also possible that some more general phenomenon is at work. Our inability to address this question in a serious way has been the result in part of the lack of a larger theoretical model such as the one proposed here. By specifying a chain of interrelated pathways that range from the macro to the micro, we can expand the scope of our investigation and identify domains of influence that have previously remained unexplored. For instance, several studies have found that social isolation is related to markers of inflammation including C-reactive protein and IL-6, though associations are more consistent in men than in women (214–218). Negative or competitive social interactions are also linked to pro-inflammatory processes (163). Below, we describe several promising frameworks where such expansion might profitably take place.

ACCELERATED AGING AND A LIFECOURSE PERSPECTIVE

We speculate that social isolation, disintegration, and disconnectedness influence mortality and therefore longevity or life expectancy in part by influencing the rate of aging of the organism. In a review on aging from a social and biomedical perspective, Berkman (219) hypothesized that social isolation "was a chronically stressful condition to which the organism responded by aging faster. Isolation would then also be associated with age-related morbidity and functional decline. Thus, the cumulative conditions [that] tend to occur in very old age [would be] accelerated." Such "accelerated aging" hypotheses have also been applied to other social experiences, especially to racial differences in health in the United States. Geronimous has developed a concept of "weathering" that describes the accelerated rate at which African Americans and other racial/ethnic minorities may age in response to harsh and discriminatory social experiences (220–225). Recent work on telomere length and allostatic load supports this perspective as well (226), suggesting

that basic "aging" processes that were once thought to be internally driven are all susceptible to stressful social and physical environments.

It is characteristic of changes related to aging that peak rises in response to stress or challenge are not as different between young and old as is the time it takes to return to prechallenge levels. Older animals take longer to return to a baseline state after challenge and therefore spend a greater amount of time "under the curve." This has implications for the cumulative wear and tear of life stressors in late life.

Missing from our earlier conceptualization in the first edition of Social Epidemiology was a lifecourse perspective, which has become much clearer over time. Research on humans and animals (both primates and non-primates) indicates that early experiences, especially social experiences between primary caregivers and infants, are powerful determinants of social, behavioral, and physiological development across the lifespan. In fact, many changes in function that are considered "normal aging" show variability related to early life experiences. It now appears that long-term neurobiological experiences which unfold in old age may have been shaped, in part, by experiences during early "critical" or "sensitive" experiences (227–231).

THE BIOLOGICAL EFFECTS OF ADULT SOCIAL EXPERIENCES: CONTINUITY AND CHANGE

Early theories of aging assumed that plasticity was a characteristic of early phases of development and was virtually nonexistent by old age. In contrast, developmental neurobiologists, neuropsychologists, social scientists, and geriatricians now recognize that in most domains, change occurs through the lifecourse and is not restricted to early development. For instance, neuronal plasticity, especially following injury, has been the subject of a great deal of research, most of which suggests that the aging brain is more plastic than we ever suspected (144, 145, 232–234). Similarly, clinical trials of physical activity across adulthood show that interventions, even in very old age, have significant effects (235–237). In fact, recent scientific initiatives related to reversing early childhood exposures is increasingly showing plasticity and resilience over the lifecourse (238–240).

The impact on health outcomes of social attachments made in early years remains an intriguing and understudied area; however, the vast body of epidemiologic evidence produced to date indicates that it is adult social circumstances that are linked to poor health outcomes. Debates in which we pitch continuity (the effect of early development/environment) against discontinuity (the effect of recent events) are not likely to be fruitful, because both have consequences for health outcomes. Furthermore, we know that large-scale social upheavals and transitions profoundly disrupt patterns of social organizations established in earlier life. Geographical relocation related to urbanization, housing policy, or employment opportunities, large-scale social change or depression such as seen in Russia and Eastern Europe, and job stress and corporate policies that are not "family friendly" represent environmental challenges that tear at the fabric of social networks, which in turn have deleterious consequences on health.

Chapter 14 of this volume discusses a range of biological mechanisms that link adult social experiences to poor health outcomes. In this chapter, we have emphasized only those that have been found to link aspects of social networks and support to health.

ASSESSMENT OF SOCIAL INTEGRATION, SOCIAL NETWORKS, AND SOCIAL SUPPORT

The assessment of aspects of social relationships in epidemiologic studies has now benefited greatly from work in the social sciences. Our aim in this section is to introduce the reader to a range of measures available, with a brief commentary regarding their utility for a specific purpose. At the outset, it should be explicit that we do not believe there is a single measure or approach that is optimal or even appropriate for all purposes. The investigator must consider why he or she hypothesizes that social ties are important to the health outcome of interest and then select and potentially modify or tailor an instrument. For instance, evidence to date suggests that measures of social integration are related to mortality and perhaps to the development of atherosclerosis, whereas emotional support is most highly related to survival in post-MI (myocardial infarction) patients. These findings and subsequent hypotheses for new studies necessitate the use of different measures. In a similar vein, studying HIV transmission or initiation of high-risk behaviors necessitates the use of still other types of instruments.

We have divided our discussion of measures into four sections: *(1)* those measures that primarily assess social ties or social integration; *(2)* measures that more formally assess aspects of social networks; *(3)* measures assessing social support, both cognitively "perceived" and behaviorally "received"; *(4)* measures of loneliness; and *(5)* measures on negative interactions and conflict. Table 7.1 shows examples of these domains along with references to the measures. The reader is referred to several lengthier reviews, especially in relation to social support.

MEASURES OF SOCIAL TIES AND INTEGRATION

Several brief measures of social ties have been used in large prospective community-based studies. They consistently predict health outcomes, particularly mortality. These scales, consisting of between 9 and 18 items, usually take between 2 and 5 minutes to administer. The instruments often tap the size of networks, frequency of contact, membership in voluntary and religious organizations, and social participation. Perhaps the best conceptual framework in which to place these measures is that of social integration. From this perspective, the measures often assess domains of social network size and diversity and social engagement and participation. Because these measures are brief, they rarely include multiple items tapping a similar domain. Therefore, with the exception of the Orth-Gomér and Johnson instrument (241), there are limited data on internal consistency from a psychometric standpoint. They do, however, have good test–retest reliability (242), are modestly correlated with other psychosocial constructs in expected ways (155, 243), and have solid construct validity in terms of consistency in predicting mortality.

The ease with which the instruments are administered, the degree to which they assess a broad range of levels of social integration from extreme isolation to high levels of integration, and their proven predictive validity are the major assets of this class of instruments. Their major disadvantages lie in not providing much insight into the mechanisms that might be health promoting (e.g., provision of emotional or instrumental support, social engagement, social influence) and

TABLE 7.1: Ways of assessing social relationships

Social Relations	
Social Network Index	(Berkman & Syme, 1979) (6)
Social Relationships and Activity	(House, Robbins, & Metzner, 1982) (8)
Social Network Interaction Index	(Orth-Gomer & Johnson, 1987) (290)
Social Contacts and Resources	(Donaldson & Ware, 1982) (242)
Social Network Assessments	
Egocentric Network Name Generators	(Antonucci, 1986 [106]; Marsden, 2005 [345]; 2006 [346]; 2011 [347])
Qualitative Network Measures	(Hollstein, 2011) (348)
Single-criterion recognition question	(Keating, Ayanian, Cleary, & Marsden, 2007) (350)
Single name generator	(Davis, Smith, & Marsden, 2007) (350)
Position generator	(Lin, Fu, & Hsung, 2001) (351)
Social Support	
Social Support Scale in OARS	(Blazer, 1982) (7)
Interpersonal Support Evaluation List (ISEL)	(Cohen & Hoberman, 1983) (248)
Social Support Scale	(Lin, Simeone, Ensel, & Kuo, 1979) (352)
Social Support Questionnaire (SSQ)	(Sarason, Levine, & Basham, 1983) (126)
Inventory of Socially Supportive Behaviors (ISSB)	(Barrera, Sandler, & Ramsay, 1981) (250)
Interview Schedule for Social Interaction (ISSI)	(Henderson, Duncan-Jones, Byrne, & Scott, 1980) (256)
Perceived Social Support (PSS)	(Procidano & Heller, 1983) (249)
Perceived Social Support Scale (PSSS)	(Blumenthal et al., 1987) (252)
Abbreviated ISSI	(Unden & Orth-Gomer, 1984) (257)
Medical Outcomes Study Social Support	(Sherbourne & Stewart, 1991) (251)
ENRICHD Social Support Inventory (ESSI)	(Mitchell et al., 2003) (255)
Survey of Children's Social Support (SOCSS)	(Dubow & Ullman, 1989) (353)
Support in Intimate Relationships Rating Scale (SIRRS)	(Dehle, Larsen, & Landers, 2001) (354)
Lubben Social Network Scale (LSNS) and Six-Item Lubben Social Network Scale (LSNS-6)	(Lubben, 1988 [355]; Lubben et al., 2006 [258])
Social Provisions Scale (SPS)	(Cutrona & Troutman, 1986) (189)
Korean Social Support Questionnaire	(Oh et al., 2008) (356)
Social Support Questionnaire	(Sarason et al., 1983) (126)

(*continued*)

TABLE 7.1: Continued

Social Support Index	(Krause & Markides, 1995) (133)
Norbeck Social Support Questionnaire (NSSQ)	(Norbeck, Lindsey, & Carrieri, 1981) (357)
Index of Sojourner Social Support (ISSS)	(Ong & Ward, 2005) (358)
Multidimensional Scale of Perceived Social Support (MSPSS)	(Zimet, Dahlem, Zimet, & Farley, 1988) (359)
Close Persons Questionnaire	(Stansfeld & Marmot, 1992) (360)
Positive Relations with Others Scale	(Ryff, 1989) (361)
Received Social Support Scale	(Vinokur, Price, & Caplan, 1996) (362)
Negative Relationships	
Positive and Negative Social Exchanges Scale (PANSE)	(Newsom, Rook, Nishishiba, Sorkin, & Mahan, 2005) (178)
Inventory of Negative Social Interactions (INSI)	(Lakey, Tardiff, & Drew, 1994) (363)
Social Undermining Scale (SUND)	(Vinokur et al., 1996) (362)
Daily Parental Withdrawal Scale	(Repetti & Wood, 1997) (364)
Negative Social Interactions at Work Scale	(Repetti, 1993) (365)
Marital Anger Scale and Marital Withdrawal Scale	(Repetti, 1989 [366]; Story & Repetti, 2006 [367])
Family Environment Scale	(Moos & Moos, 1981) (368)
Dyadic Adjustment Scale	(Spanier, 1976) (369)
Loneliness	
Three-Item Loneliness Scale	(Hughes, Waite, Hawkley, & Cacioppo, 2004) (263)
Revised UCLA Loneliness Scale	(Russell, Peplau, & Cutrona, 1980) (370)
De Jong Gierveld Loneliness Scale	(de Jong Gierveld & van Tilburg, 1999) (371)
Emotional/Social Loneliness Inventory	(Vinconzi & Grabosky, 1987) (372)
Social and Emotional Loneliness Scale	(DiTommaso, Brannen, & Best, 2004) (373)
Loneliness and Social Dissatisfaction Questionnaire (LSDQ)	(Asher, Hymel, & Renshaw, 1984) (374)
Philadelphia Geriatric Morale Scale on Lonely Dissatisfaction	(Lawton, 1975) (375)
Paloutzian and Ellison Loneliness Scale	(Paloutzian & Ellison, 1982) (376)
Older Americans Resources and Services (OARS) Social Resource Rating Scale regarding loneliness frequency	(Duke University, 1978 [377]; Morrow-Howell, Becker-Kemppainen, & Judy, 1998 [378])
Worker Loneliness Questionnaire	(Chadsey-Rusch, DeStefano, O'Reilly, Gonzalez, & Collet-Klingenberg, 1992) (379)

in providing limited information on the depth and quality of social relationships. Since it is likely that the critical mechanisms vary among health outcomes, this can be a serious shortcoming of the measures.

ASSESSMENT AND MEASURES OF SOCIAL NETWORKS

Most classical measures of social networks have been developed without an eye toward how they might be used in studies of health outcomes. However, they provide the best measures of network structure and are often sensitively linked to aspects of social support, and occasionally to patterns of social influence or person-to-person contact that enable transmission of infectious agents. Most instruments take between 20 minutes to an hour to complete and provide a rich understanding of the complex dynamics and morphology of networks. First-generation classical examples are those developed by Fischer (40), Wellman (37), and Laumann (36). In the past decade, several network instruments have been adapted from these earlier assessments with modifications for use in epidemiology and health psychology. Antonucci's (106) convoy measure makes excellent use of a bull's-eye mapping technique used in traditional network assessment (244). Following the network assessment, the subject provides information about individuals in the network regarding social support and sociodemographic characteristics. Similarly, measures from our group, based on the Yale Health and Aging Study, were adapted from questions developed by Fischer in his California study (40). The items tap critical dimensions of networks (size, homogeneity, density, contact, proximity) and support (types, availability, adequacy, source) in an abbreviated fashion without asking to identify specific individuals (see 155, 245). These measures are not as lengthy to administer as the traditional network questionnaires but they are also not as rich in assessment of the full range of characteristics as are the traditional measures.

However, over the last decade formal social network analysis with extensive egocentric and sociocentric modeling have been developed and successfully used in studies linking network structure and function to health and health behaviors with critically important results. If the aim of an investigator is to test hypotheses related to specific structural components of networks (e.g., homogeneity, multiplexity, density, reachability), these instruments are ideal and should be used more often in health-related research. Next we come to identifying critical domains of social networks. A social network might be defined as the web of social relationships that surround an individual and the characteristics of those ties (23, 36, 39, 40, 246). Burt has defined network models as describing "the structure of one or more networks of relations within a system of actors" (247). In this chapter, we consider both egocentric networks (networks surrounding an individual) and sociocentric approaches of full networks more commonly incorporated in network analyses. Network characteristics cover nodes, ties, and networks:

Characteristics of Nodes

- Reachability (existence of a path between two nodes)
- Structural equivalence (having the same types of ties to the same people)
- Average popularity (number of connections held by one node relative to other nodes in the network)

- Centrality (number of nominated paths received by a node relative to the number of nominated paths suggested by the node)

Characteristics of Ties

- Timing (when ties form) and duration (the length of time an individual knows another)
- Multiplexity (the number of types of transactions or support flowing through a set of ties)
- Distance measures (the length of a path between two nodes)
- Frequency of contact (number of face-to-face contacts and/or contacts by phone or mail)
- Reciprocity (existence of a mutual path between nodes)
- Transitivity (existence of clustering or triangular relations between nodes)
- Bridging (existence of a path between a node in one group and a node in another group)

Characteristics of Networks

- Range or size (number of network members)
- Boundedness (the degree to which they are defined on the basis of traditional group structures such as kin, work, neighborhood)
- Density (the extent to which the members are connected to each other)
- Structural cohesion (number of independent paths that hold together group members)
- Clustering and segregation (presence of groups within networks)
- Homogeneity (the extent to which individuals are similar to each other in a network)
- Roles, relations, and block modeling (pattern of roles underlying relations in the network)
- Stability (frequency with which network membership changes)

Several software packages have been developed for drawing and analyzing social networks. The broadening of social network analysis from egocentric analysis focusing on dyadic ties between two individuals to larger, more complex analyses of superdyadic networks has required and facilitated technological innovation in network analysis software programs. Table 7.2 shows the approach, purpose, and resources for using this new network approach.

TABLE 7.2: Methods for social network analysis

Approach	Purpose	Resources and references
Social Network Image Animator (SoNIA)	Program that visualizes dynamic networks that may change over time. SoNIA can be used with PAJEK, UCINET, or with various packages from R.	(Moody, McFarland, & Bender deMoll, 2005) (153) http://sonia.stanford.edu/
PAJEK	Network drawing and analysis software. PAJEK is ideal for use with large networks. It is also possible to link PAJEK with R.	(Batagelj, Mrvar, & de Nooy, 2008) (380) (Batagelj & Mrvar, 2001) (381)

(*continued*)

TABLE 7.2: Continued

Approach	Purpose	Resources and references
NetMiner	Network visualization software. NetMiner is best for use with smaller networks.	(Cyram, 2004) (382)
NetDraw	Network visualization software containing multidimensional scaling (MDS) techniques. NetDraw is best for use with smaller networks.	(Borgatti, 2002) (383) https://sites.google.com/site/netdrawsoftware/
Krackplot	Network visualization software containing multidimensional scaling (MDS) techniques.	http://www.andrew.cmu.edu/user/krack/krackplot.shtml
UCINET	Network analysis program that is compatible with PAJEK and NetDraw. UCINET is not ideal for large networks.	(Borgatti, Everett, & Freeman, 1999) (384) https://sites.google.com/site/ucinetsoftware/home
NEGOPY	One of the first social network analysis software programs. It is best for conducting subgroup analyses.	http://www.sfu.ca/personal/archives/richards/Pages/negopy.htm
R	Contains several packages that allow people to conduct social network analyses. These include statnet or ergm, a package that works within statnet. Both of these use exponential-family random graph models. Statnet allows for model estimation, evaluation, simulation, and visualization. Ergm works within the statnet package, enabling users to conduct model simulations and visualizations as well as goodness of fit tests.	**R and related resources are available at:** http://www.r-project.org/ **Statnet and related resources:** http://statnet.csde.washington.edu/ http://cran.r-project.org/web/packages/statnet/index.html (Goodreau, Handcock, Hunter, Butts, & Morris, 2008) (385) (Handcock, Hunter, Butts, Goodreau, & Morris, 2008) (386) **Ergm and related resources:** http://cran.r-project.org/web/packages/ergm/index.html http://cran.r-project.org/web/packages/ergm/ergm.pdf (Hunter, Handcock, Butts, Goodreau, & Morris, 2008) (387)
SAS Programs for Analyzing Networks (SPAN)	Set of programs that interface with other social network analysis programs. SPAN can be used to draw networks, transfer network data from and to other analysis programs, calculate network measures, and analyze models. It can be used with large networks and can handle multiple networks simultaneously.	http://www.soc.duke.edu/~jmoody77/span/span.zip

MEASURES OF SOCIAL SUPPORT

Over the last 15 years, there has been a proliferation of social support measures. They often share a core set of orientations, particularly in the assessment of several types of support including emotional, instrumental or tangible, appraisal, and financial. Beyond that core, the measures are often different from one another in subtle yet important ways. Perhaps the most striking difference is in the orientation to the assessment of perceived support versus received support. For instance, perceived support items are often oriented toward hypothetical conditions ("If you need help, is there anyone you could count on for a small loan, or help with a problem?"). Received support is often grounded in behavioral transactions occurring over a set period of time ("In the last week, month, etc., did anyone talk to you about your feelings, lend you money?"). The investigator must choose between these orientations depending on the hypothesis and population being studied.

Social support instruments tend to have been better studied with regard to their psychometric status. Furthermore, since they commonly include multiple items tapping single domains, they have good internal validity. They usually include from 15 to 40 items and take between 10 and 20 minutes to administer. Their only weakness from the perspective of external validity is that they have often been developed on a very small, typically college-aged population. Their applicability to populations of heterogeneous middle-aged and older adults must be ascertained on a case-by-case basis. It should be noted that pure social support instruments such as those developed by Cohen (248), Procidano and Heller (249), Barrera (250), Sherbourne and Stewart (251), Blumenthal et al. (252–255), and Sarason et al. (126) are excellent measures of support but do not measure network structure (and do not purport to do so). If the investigator is interested in a specific aspect of social support, these are excellent choices for use and ease of administration.

Early in the development of assessments of social interaction, Henderson (256) developed an excellent measure encompassing a broad range of dimensions including social integration, social interactions, and attachment. This instrument has been used in a range of settings primarily with regard to psychiatric status. It has 52 items and takes about 30 minutes to complete. It has been modified by Unden and Orth-Gomér (257) to take under 10 minutes, and is very useful in covering a range of dimensions not exclusively falling into any single domain.

MEASURES OF LONELINESS

The assessment of loneliness in contrast to network assessments or measures of social integration and engagement rests not so much on behaviors as it does on perceptions of feeling lonely. As noted by Hawkley and Cacioppo, loneliness is defined as a "distressing feeling that accompanies the perception that one's social needs are not being met by the quantity or especially the quality of one's social relationships" (17). Loneliness measures draw on perceptions rather than actual behaviors. Measures of loneliness are often based on an instrument developed by Lubben called the UCLA Loneliness Scale (258). It includes items such as feeling isolated, feeling part of a group, having people one can talk to. Sometimes individual items from the CES-D are used (e.g., "I feel lonely"). Hawkley et al. describe chronic loneliness as experienced by 15–30% if the population, although they acknowledge that we all feel lonely from

time to time (17). In a series of investigations, Cacioppo and colleagues reported that while loneliness is related to network structure in various ways (14), it has independent effects on health outcomes, psychophysiologic reactions, and health behaviors (12–14, 17, 18, 66, 67, 170–172, 259–265).

MEASURES OF NEGATIVE SOCIAL INTERACTIONS

As measures of social interactions have become more nuanced, increasingly there are scales developed to assess negative interactions. These include conflict, demands, and abuse. Negative partner interactions have been associated with elevated hypothalamic-pituitary-adrenal (HPA) axis responses (266) and more recently with allostatic load (267). Negative aspects of domains—sometimes called social negativity (268)—have been increasingly recognized as influencing health outcomes. More recently, measures developed by the MacArthur Mid-Life in the US (MIDUS) network have been included not only in the MIDUS Study but also in the Health and Retirement Study. Negativity questions from MIDUS are asked of a spouse/partner (if applicable), friends, and nonpartner family. Social negativity is measured from each of these sources by asking respondents how often each source "makes too many demands on you," "criticizes you," "lets you down when you are counting on him/her," and "gets on your nerves." For spouse and partner, additional items are asked about arguing and making you feel tense (162, 266, 269, 270). Measures have a Cronbach's alpha of 0.77, 0.78, and 0.87 for family, friend, and partner scales, respectively.

Our aim in this section has been to give the reader a brief overview of the spectrum of measures available to assess aspects of social networks and relationships. There are many instruments in this area now, and our goal was not to be comprehensive, but, on the contrary, to identify instruments we believe have a great deal of promise and utility to epidemiologists.

SOCIAL NETWORKS AND MORTALITY, MORBIDITY, FUNCTION, AND BEHAVIOR

Beginning in the 1970s and continuing up until the present, a vast literature has accumulated that links social networks or social support to mortality, morbidity, function, and behavioral risks. A complete review of this literature is beyond the scope of this chapter, and the reader is referred to several recent reviews covering a broad array of outcomes (9, 10, 23, 271–275). Because of the vast literature on mental health and numerous recent reviews (13, 17, 18, 60, 144, see 145, 276–278), we have not included this in our discussion in any lengthy way, except with regard to cognition. Our intention here is to review the evidence linking social networks and social support to selected outcomes, highlighting studies related to all-cause mortality, cardiovascular disease, stroke, and infectious diseases. We are also interested in the mediating pathways leading to poor health, especially behavioral risks related to tobacco consumption, drug use, contraceptive practices, and participation in prevention programs.

ALL-CAUSE MORTALITY

Over the last 30–35 years, well over 100 studies have been conducted examining the effects of social networks and relationships on mortality risk. In our original edition of this book, we cited 13 large prospective cohort studies across a number of countries from the United States to Scandinavian countries to Japan. A recent meta-analysis (9) conducted in 2010—about 30 years after the initial Alameda County findings were reported—identified 148 studies that had the capacity to contribute specific statistical information to a meta-analysis. This represents over a tenfold increase since the first edition of our book. These studies have been conducted in a number of countries from Australia, Japan, and Israel to countries in North and South America and Europe. Still, very few studies have come from low-income countries. A recent analysis of self-reported health and social ties, however, suggests that the association between social relationships and health is nearly universal (279). The studies linked to mortality include general populations from large, longitudinal cohorts to highly specialized cohorts of men and women with specific diseases and conditions in which case-fatality is the major issue. The meta-analysis reveals that, on the whole, people who are isolated or disconnected from others are at increased risk of dying prematurely. The results suggest that the odds ratios (ORs) for isolation are about 1.5, indicating a 50% increased overall risk of dying. When studies of either structural or functional aspects of networks were analyzed separately, the ORs ranged from 1.57 for structural to 1.46 for functional aspects. We interpret these as very small differences from the overall evaluations, suggesting that both are important and may in fact be highly correlated. In fact, complex measures with more components (and items) had the highest ORs, suggesting the benefit of psychometrically sound multicomponent models.

Here, we briefly review the initial studies with historical value in this field, and refer the reader to the fuller review in the meta-analysis, as well as reviews by Christakis (23) and Seeman (60, 274) referenced above. In the first of these studies—from Alameda County (6)—men and women who lacked ties to others (in this case, based on an index assessing contacts with friends and relatives, marital status, and church and group membership) were 1.9 to 3.1 times more likely to die in a 9-year follow-up period from 1965 to 1974 than those who had many more contacts. The relative risks associated with social isolation were not centered on one cause of death; rather, those who lacked social ties were at increased risk of dying from ischemic heart disease (IHD), cerebrovascular and circulatory disease, cancer, and other causes in a final category that included respiratory, gastrointestinal, and all other causes of death. Clearly, this social condition is not associated exclusively with increased risks from, say, coronary heart disease (CHD). The relationship between social isolation and mortality risk was independent of health behaviors such as smoking, alcohol consumption, physical activity, preventive health care, and a range of baseline comorbid conditions.

Another study—this one in Tecumseh, Michigan (8)—shows a similar strength of positive association for men, but not for women, between social connectedness/social participation and mortality risk over a 10- to 12-year period. An additional strength of this study was the ability to control for some biomedical predictors assessed by physical examination (e.g., cholesterol, blood pressure, and respiratory function). In the same year, Blazer (7) reported similar results from an elderly sample of men and women in Durham County, North Carolina. He compared three measures of social support and attachment: *(1)* self-perceived impaired social support, including feelings of loneliness; *(2)* impaired social roles and attachments; and *(3)* low frequency of social

interaction. The relative risks for dying associated with these three measures were 3.4, 2.0, and 1.9, respectively.

In the late 1980s and 1990s, results from more studies were reported—one from a study in the United States and three from Scandinavia. Using data from Evans County, Georgia, Schoenbach et al. (280) used a measure of social contacts modified from the Alameda County Study and found risks to be significant in older white men and women even when controlling for biomedical and sociodemographic risk factors, although some racial and gender differences were observed. In Sweden, the Göteborg Study (281) shows that in different cohorts of men born in 1913 and 1923, social isolation proved to be a risk factor for dying, independent of age and biomedical risk factors. A report by Orth-Gomér and Johnson (241) is the only study besides the Alameda County one to report significantly increased risks for women who have been socially isolated. Finally, in a study of 13,301 men and women in eastern Finland, Kaplan and associates (282) have shown that an index of social connections almost identical to the Social Network Index used in Alameda County predicts mortality risk for men but not for women, independent of standard cardiovascular risk factors.

Several studies of older men and women in the Alameda County study and the Established Populations for the Epidemiologic Study of the Elderly (EPESE) Studies confirm the continued importance of these relationships into late life (155, 156). Furthermore, two studies of a large cohort of men and women in a large health maintenance organization (HMO) (283) and 32,000 male health professionals (284) suggest that social networks are, in general, more strongly related to mortality than to the incidence or onset of disease.

Two studies of Danish men (285) and Japanese men and women (286) further indicate that aspects of social isolation or social support are related to mortality. A more recent study in France of French Gas and Electricity employees finds similar overall mortality risks, though risks are elevated for cancer mortality and not for cardiovascular disease mortality (287). Virtually all of these studies find that people who are socially isolated or disconnected to others have between two and five times the risk of dying from all causes compared to those who maintain strong ties to friends, family, and community.

CARDIOVASCULAR DISEASE

There is conflicting, albeit limited, evidence that social networks or support are related to the onset of cardiovascular disease. One study of middle-aged Swedish men shows social integration to be related to the incidence of MI (288), but several other studies have reported no associations (283, 284).

In contrast, in the last 6 years, there have been a host of studies suggesting that social ties, especially intimate ties and emotional support provided by those ties, influence survival among people post MI or with serious cardiovascular disease. In the first of these, Ruberman et al. (289) explored 2,320 male survivors of acute MI who were participants in the Beta-Blocker Heart Attack Trial. Patients who were socially isolated were more than twice as likely to die over a 3-year period than those who were less socially isolated. When this measure of social isolation was combined with a general measure of life stress, which included items related to occupational status, divorce, exposure to violent events, retirement, or financial difficulty, the risks associated with high-risk

psychosocial status were even greater. Those in the high-risk psychosocial categories were four to five times as likely to die as those in the lowest risk categories. This psychosocial characteristic was associated with death from all causes and sudden deaths. It made large contributions to mortality risk in both the high-arrhythmia and low-arrhythmia groups. In this study (and in most of the studies in which subjects are recruited post event), the investigators were not able to determine the temporal association between the assessment of psychosocial resources and the severity of disease. Nonetheless, it serves as a powerful model for future studies.

In a second Swedish study of 150 cardiac patients and patients with high-risk factor levels for CHD, the finding that lack of support predicts death was further confirmed (290). Patients who were socially isolated had a 10-year mortality rate that was three times higher than did those who were socially active and integrated. Because these patients were examined extensively for prognostic factors at study entry, it was possible to disentangle effects of psychosocial and clinical characteristics.

In a third study, Williams et al. (291) enrolled 1,368 patients undergoing cardiac catheterization from 1974 through 1980 who had been found to have significant coronary artery disease. They examined survival time until cardiovascular death through 1989. In this study, men and women who were unmarried or without a confidant were over three times as likely to die within 5 years compared with those who had a close confidant or who were married (OR: 3.34; 95% confidence interval [CI]: 1.8, 6.2). This association was independent of other clinical prognostic indicators and sociodemographic factors, including socioeconomic status.

Case et al. (292) examined the association between marital status and recurrent major cardiac events among patients post MI who were enrolled in the placebo arm of a clinical trial, the Multicenter Diliazem Post-Infarction Trial. These investigators reported that living alone was an independent risk factor, with a hazard ratio of 1.54 (95% CI: 1.04, 2.29) for recurrent major cardiac event, including both nonfatal infarctions and cardiac deaths.

In a fifth study, we explored the relationship between social networks and support and mortality among men and women hospitalized for MI between 1982 and 1988 who were participants in the population-based New Haven EPESE (293). Over the study period, 100 men and 94 women were hospitalized for an MI. Thirty-four percent of women and 44% of men died in the 6-month period after MI.

Among both men and women, emotional support—measured prospectively—was related to both early in-hospital death and later death over a 1-year period. Among those admitted to the hospital, nearly 38% of those who reported no source of emotional support died in the hospital, compared with 11.5% of those with two or more sources of support. The patterns remained steady throughout the follow-up period. At 6 months—the major end point of the study—52.8% of those with no source of support had died, compared with 36.0% of those with one source and 23.1% of those with two or more sources of support. These figures did not change substantially at 1 year. As Figure 7.2 shows, the patterns were remarkably consistent for both men and women, younger and older people, and those with more or less severe cardiovascular disease, as assessed by a Killip classification system. In multivariate models that controlled for sociodemographic factors, and psychosocial factors including living arrangements, depressive symptoms, and clinical prognostic indicators, men and women who reported no emotional support had almost three times the mortality risk compared with subjects who reported at least one source of support (OR: 2.9, 95% CI: 1.2, 6.9).

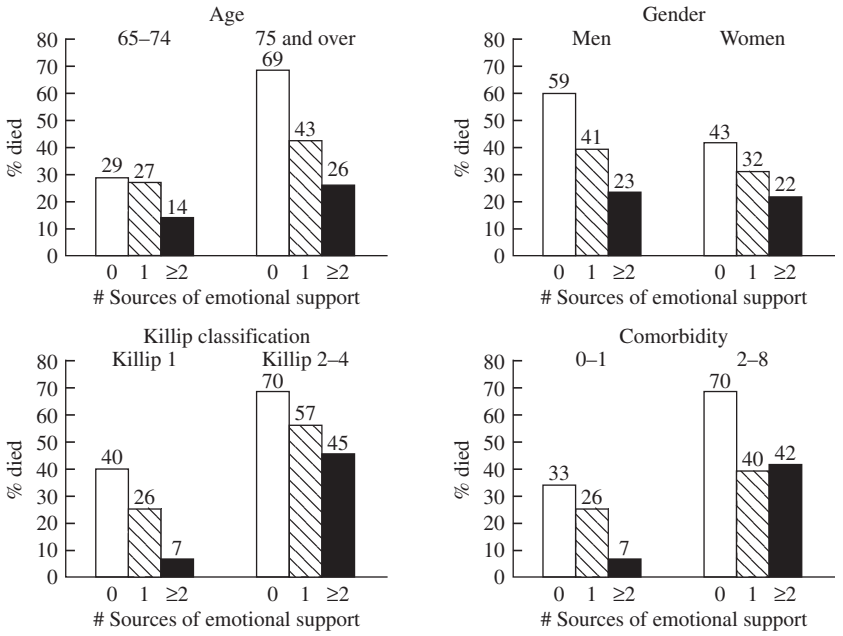


FIGURE 7.2: Percentage of patients with myocardial infarction who died within 6 months ranked by level of social support. Adjustments were made for age (top left), gender (top right), severity of myocardial infarction as defined by Killip class (bottom left), and comorbidity (bottom right).

Source: Adapted from Berkman, Leo-Summers, and Horowitz, 1992 (293).

In a study of men and women undergoing coronary bypass surgery or aortic valve replacement, Oxman and colleagues (294) found that membership in voluntary organizations, including religious organizations, and drawing strength and comfort from religious or spiritual faith were related to survival post surgery. When these two dimensions were combined, people who endorsed neither of these items were over seven times as likely to die as those who belonged to such organizations and drew comfort from their faith. Though it is beyond the scope of this chapter to go into detail on the recent research on religiosity, this later study complements and balances the work on the importance of intimacy by illustrating that a sense of belonging to informal organizations that are rooted in common values and collective goals may also be an important influence on well-being and survival.

In a study of Mexican Americans and non-Hispanic whites in the Corpus Christi Heart Project (295), social support was found to predict mortality for an average period of over 3 years; however, the relative risk was very strong in the Mexican American men and women (RR: 3.38, 95% CI: 1.73, 6.62).

These findings in post-MI populations, coupled with the strong data on long-term mortality and relatively weaker data on incidence, would suggest that social networks and support may have the greatest impact on determining not the onset of disease but rather prognosis and survival.

To date there are only a handful of studies related to other cardiovascular related diseases. A study of congestive heart failure (296) among older men and women in New Haven found emotional support to be related to survival for men but not women and found no association with risk for initial hospitalization (297).

STROKE

As we noted with respect to cardiovascular disease, the evidence in favor of the view that social integration is associated with cerebrovascular disease is less compelling for incidence and to some extent for mortality. However, the evidence that social networks and support are important in recovery from stroke is increasingly convincing.

Several studies have identified a trend toward higher risk of death from stroke among those who are socially isolated (243, 284), although these studies have lacked the strength to evaluate the associations fully. However, a number of additional studies have shown that social networks and support (particularly social isolation) are associated with case fatality in the post-stroke period. For example, in a study by Vogt et al. (283), social network measures were strong predictors of both cause-specific and all-cause mortality among persons who had incident cases of ischemic heart disease, cancer, and stroke. During 10 years of follow-up of a group of newly diagnosed stroke patients, clinical diagnosis of depression was associated with poor survival (298). In that study, patients who were both socially isolated and clinically depressed were at particular risk for post-stroke fatality. To date, no studies have reported a link between social isolation and incidence of nonfatal stroke. In one study of 32,624 US male healthcare workers, Kawachi found a trend in the association between risk of nonfatal stroke and social networks. However, it was not possible to conduct multivariate analyses due to inadequate statistical power (284). What seems clear is that the evidence in favor of a link between social ties and disease incidence has been shown only for certain infectious diseases and, to a limited extent, for coronary heart disease. Efforts to identify an association in stroke have suffered from inadequate statistical power. In theory, the same mechanisms that are likely to be associated with protection against heart disease may operate in stroke, although they may be more difficult to detect when the number of events becomes small. Chief among them may be modulation of blood pressure (299) and stress-related vascular reactivity.

Numerous observational studies have reported that several aspects of social integration, particularly operating through emotional support, influence stroke recovery both in terms of physical functioning and psychological adjustment (203, 300–304). Several studies have found that social support predicts quality of life after stroke (305–308). The absence of social support has been shown to be associated with a variety of negative responses to stroke including suicidal thoughts (309) and post-stroke depression (PSD) (271). The availability of social support has also been shown to be an important predictor of hospital course, including length of stay and discharge disposition (310, 311). In a cohort study of 152 stroke survivors, Brosseau and colleagues (312) found that the presence of social support predicted both discharges to rehabilitation and discharges to nursing homes. The findings regarding the impact of social support on stroke recovery appear to be particularly robust in light of a recent review that discards those studies that failed to adhere to sound methodological principles (313). In that study, social support was the only psychosocial factor to be retained.

In addition, several randomized clinical trials have shown that psychosocial interventions have led to improved adjustment in stroke patients (314) and longer survival in patients with other chronic illness. Enhancement of available social supports has been an important element in these intervention approaches (315, 316).

COGNITIVE FUNCTION: RESERVE AND RESILIENCE

There is growing evidence that social networks—particularly those related to social engagement and interaction—influence cognitive function in adulthood. Social engagement and network activities may be most important in relation to cognitive function; they may reduce the rates of decline through several mechanisms including the cognitive demands of social interaction such as receptive and expressive communication, recall of experiences, and problem solving, which may have direct effects for neurologic functioning (144). Network members may also encourage health-promoting behaviors or provide direct care and access to medical care. Increasingly, cognitive scientists distinguish between cognitive reserve, a condition producing good cognition developing over the lifecourse, and cognitive resilience, a phenomenon that helps to restore cognition after challenges. Most of the evidence to date suggests that active social engagement and ties with others may be most important in producing cognitive reserve. Findings indicating the importance of social support are not as strong with regard to reserve, although support may help modify outcomes after stressful experiences such as strokes in helping with cognitive resilience.

Over the last 25 years, a number of studies have shown that being socially engaged has been associated with risks of cognitive decline among older adults. One of the first of these studies was in the New Haven cohort of the EPESE studies. Bassuk et al. (317) show that, compared with men and women with many ties, those with no social ties were at increased risk of incident cognitive decline in this cohort of 2,812 men and women interviewed over a 12-year period. The odds ratios of having an incident decline over this period—in which four waves of longitudinal data were collected—were 2.37 (95% CI: 1.07, 4.88). These results were seen when controlling for a host of covariates including age, education, and physical function, among others. Studies in Sweden focusing on onset of dementia further implicated the important role of social engagement in preventing dementia onset (318, 319). More recently, results from the Kungsholmen study suggest that mental, physical, and social aspects of activities each contribute to dementia risks. The most protective experiences are those that combine 2 of the 3 aspects (320). In the Chicago Health and Aging Project, Barnes et al. (321) also report that both social networks (assessed as number of ties) and social engagement (assessed from social activities) were associated with cognitive decline. In another study of older Spanish men and women, Zunzunegui et al. (278) also report that visual contact with relatives and community social integration are associated with declines in cognitive function, while engagement with friends may be more important for women than for men. In a study of elders in Taiwan, Glei et al. (322) find that, rather than the number of close ties with friends and relatives, social activities such as volunteering or socializing with friends are associated with cognitive decline. They note that in a tightly knit family structure such as the one in which most elderly Taiwanese find themselves, it may be activities, rather than less voluntary but more intimate networks, that are protective. The authors note that network ties in Taiwan may also reflect needs for care, so that networks—as well as the protective effects of social relations and engagement—may be stronger for those who are sick (reflecting the impact of frailty on network support).

Although most of the studies discussed so far suggest that social engagement precedes cognitive declines, some evidence suggests that this may not be true. Findings from the Honolulu Heart

Study report that midlife measures of social engagement do not predict incident dementia at older ages, and that, in fact, only declines in engagement from midlife to old age are associated with dementia onset (323). These findings would suggest that disengagement might be a prodromal sign of early onset. The issues of reverse causation are complex; this study is important in its establishment of temporality: measurements of social engagement were made in midlife, years before the likely onset of dementia or cognitive declines.

Memory seems to be particularly related to social ties and engagement. In an analysis of the Health and Retirement Study, a longitudinal panel study of Americans ages 50 and over, Ertel et al. show that social engagement is related to memory and memory declines (324). Bosma et al. (325) find that memory, verbal fluency, and executive function are associated with social engagement. Hultsch et al. (326) also show changes in other cognitive domains including executive function may be influenced by social patterns of engagement.

This work coincides with the research on the role of retirement and volunteer activities, both of which further suggest that work (whether paid or unpaid) requires maintenance of cognitive skills as essential job functions and that these roles are then central to maintaining function later in life. In a pilot study of the Experience Corps randomized study, Carlson et al. (327) examined whether they could identify brain plasticity in those who participated in the program. Such brain plasticity assessed from fMRI scans is thought to be related to executive function, which in turn is linked to both memory and functional difficulties. Those who were randomized to the Experience Corps exhibited more brain plasticity in response to this environmental enrichment. As quoted in their work, one of the participants noted, "it [Experience Corps] removed the cobwebs from my brain" (327). There is a larger literature on this related to working at older ages. This area is reviewed in depth in Chapter 5, which covers working conditions and health.

In contrast to the findings from studies on social networks, engagement, and integration, in a study of stroke recovery, emotional support emerged as the most important predictor of cognitive recovery assessed via a battery of seven neuropsychological exams (144). While social network ties were associated with cross-sectional cognitive domains, only emotional (and not instrumental) support predicted recovery over 6 months among stroke patients. These results suggest that emotional support may promote cognitive resilience, while social networks may provide cognitive reserve important to protect against initial insults after events such as stroke (144). This buffering effect of support is also evident in recent MIDUS findings in which social support was seen to buffer the effects of social strain (negative social interactions) on complex cognitive tasks related to executive function (146). Thus, as Tun et al. note, "even when individuals are faced with a very stressful environment, the opportunity for supportive, caring interactions with others may serve as a buffer for cognitive function."

INFECTIOUS DISEASE: THE ORIGINS OF SOCIAL NETWORK EPIDEMIOLOGY

There is no other set of diseases better suited to network analysis that those that are spread from person to person. Thus, sexually transmitted diseases and illnesses spread through directly shared behaviors (condom use, needle sharing, etc.) have the potential to be deeply understood

by applying network methods. Perhaps the most sophisticated and enlightening work on social networks and health has evolved from the work on HIV/AIDS. In fact, HIV/AIDS scientists have invoked the term "network epidemiology" to describe this approach. These approaches, so useful for the spread of other sexually transmitted diseases, have now been well developed in work on adolescent networks with the spread of sexually transmitted diseases and even related to information about contraceptive and sexual behaviors. The centrality of social networks in the distribution of infectious disease is succinctly described by Morris (147) in an overview done almost 20 years ago in which she argued that:

> Infectious diseases spread by person-to-person contact may be strongly channeled by patterns of selective (or "non-random") social mixing. The more intimate and extended the contact needed for disease transmission, the more impact selective mixing will have on the speed and direction of spread. Patterns of selective mixing at the population level are in turn the outcome of the heterogeneity in individual contact networks. (147)

The concern with social networks as a factor in the health status of populations is precisely the concern with the nature of these patterns of social mixing. As such, infectious diseases offer a strategic site to study several important pathways through which social network structure impacts health.

Studies of social networks and infectious disease have clustered in several discernible areas. In this selective review, we will first highlight the very important work that has been conducted in relation to HIV/AIDS over the last two decades. This work relates to other sexually transmitted diseases as well. Second, we will examine important new evidence showing the influence of social support on susceptibility to infection by the common cold virus.

SOCIAL NETWORKS AND HIV/AIDS: NETWORK EPIDEMIOLOGY ACROSS THE GLOBE

Network analysis was initially used in HIV/AIDS work to identify egocentric patterns of sexual networks. Now some of the most formal mathematical modeling in network analysis is done in relation to AIDS research. Recent work uses both egocentric and sociocentric approaches and draws on both observational studies and simulations of potential patterns of transmission, which may shape the AIDS epidemic into the future. These models make a major contribution to existing classical epidemiological mathematical models of infectious disease transmission that rely on assumptions about disease transmission that are far more random than are likely to exist in the world of sexual networks. Several key studies have used sociocentric approaches in countries in which the AIDS epidemic is strong. For instance, the Likoma Network Study (LNS) focuses on an island in sub-Saharan Africa in which several villages are surveyed (75, 79, 97). Figure 7.3 shows network structure in Likoma within three years of the survey (a), within one year of the survey (b), and at the time of the survey (c).

Social network analyses have also been conducted in Malawi using data from the Malawi Diffusion and Ideational Change Project to investigate the importance of social interactions to perceived vulnerability to AIDS (100).

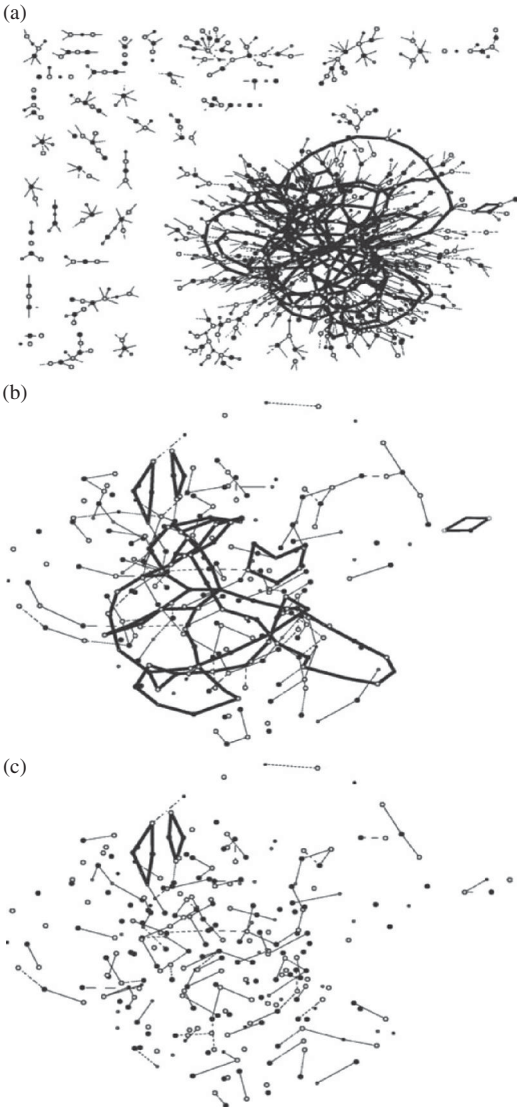
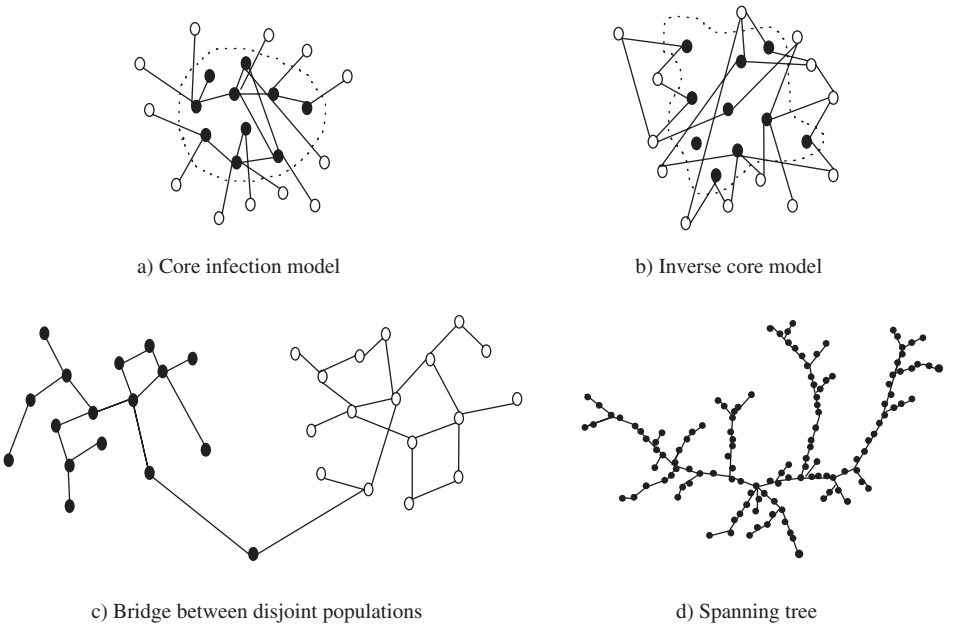


FIGURE 7.3: Visualizations of network structure from the Likoma Island Study.

Source: Helleringer and Kohler, 2007 (79).

THE CHANGING VIEW OF ROMANTIC AND SEXUAL NETWORKS AND SEXUALLY TRANSMITTED DISEASES

Until very recently, epidemiologists tended to think of risky sexual networks as containing a few members with a large number of partners or a "core" of high-activity actors sexually active within a densely knit group with some links to outside members which diffused disease. Epidemiologic models of infectious disease transmission also commonly assumed random mixing for a contact structure. But sex is not a sneeze, and is rarely transacted on a bus. These common assumptions



a) Core infection model

b) Inverse core model

c) Bridge between disjoint populations

d) Spanning tree

FIGURE 7.4: Four types of sexual networks.
Source: Bearman, Moody, and Stovel, 2004 (81).

turn out to be incorrect in describing sexual networks and the sexual transmission of many diseases, especially HIV/AIDS.

Romantic and sexual networks share many of the characteristics of broader social networks but are qualitatively and quantitatively different. Sexual networks directly influence transmission of disease through sexual contact, but they also play a role in the diffusion of safe sex practices. Bearman, Moody, and Stovel (81) describe four stylized images of sexual networks: *(1)* a core in which a group of high-activity actors pass infection to one another and diffuse infection out to a less connected population (see Figure 7.4); *(2)* an inverse core, which is a central group such as truckers infected with HIV who spread disease out to others but not directly to themselves; *(3)* bridging networks where two groups engaged in different behaviors have network members who bridge and link entire groups to each other; and *(4)* a spanning tree, which is a long chain of connections that stretch across a population with very sparse density. Bearman, Moody, and Stovel (81) note that "spanning tree structures appear when. . . rules preclude the enactment of specific relations."

Results from a sociocentric network analysis of romantic and sexual networks in 832 high school students from one high school in the Add Health Study show that a spanning tree structure best describes the sexual networks of adolescents in this high school. Most surprising is that over half the students in the school are linked in one very large spanning tree. While most students are involved in dyads or triads, in less than two years more than 50% of students are chained together via sexual or romantic networks. Sexually transmitted disease risk is not a matter of number of partners but rather of being part of a large sexual network in which most members have no way of knowing they are involved. From the perspective of disease transmission, this is a very risky situation, since most members only report one or two ties, yet disease can spread very efficiently.

Consistent with this finding and representing a clear debunking of the "high-activity" hubs is the recent research on network characteristics that might explain racial/ethnic disparities in sexually transmitted diseases, particularly between African Americans and whites in the United States. Non-Hispanic blacks have higher rates of HIV as well as of most other sexually transmitted diseases based on multiple surveys (76). While some of this increased prevalence may be related to limited medical care access, treatment, and care, it is likely that patterns of transmission account for some part of this increased risk. Early assumptions were based on superconnective "cores," as described above; however, recent simulations suggest that low-degree concurrency—exactly as described in the spanning tree model by Bearman et al. (81)—coupled with high levels of network segregation, may produce prevalence patterns seen in US surveys (76). It turns out that small differences in the configuration of partnerships from small changes in concurrent relations to smaller changes in monogamous relations have large effects on epidemic potential. When coupled with race/ethnic baseline infection rates, such patterns do not have to involve highly active cores to explain patterns of disease. Increases in numbers of partners, whether concurrently or sequentially, is related to risk of infection. However, the fraction of infected partners who transmit infection is related to both cumulative partners and concurrency.

The spread of HIV/AIDS in Africa, especially in sub-Saharan Africa, has been of major interest, as the prevalence of AIDs increases even if incidence drops in many areas. Several valuable studies have been launched over the last decade, notably in Malawi (Likoma Network Study) (75), Thailand (328), and South Africa (329–331). The Likoma Network Study serves as an excellent model of sociocentric network studies in sub-Saharan Africa; because the design is so useful, we describe it here. Likoma Island is a small island on Lake Malawi, enabling network analysts to do "small world" studies of complete networks in selected study villages. Strategies for defining sexual partners and for identifying them have been described in detail (75).

Research on the social networks of injection drug users illustrates the myth that social networks and support are inevitably health promoting (332). Some evidence suggests that the overall density of risk-taking subnetworks is associated with higher levels of risk for the individual. Participation in risk networks socializes an individual to a health-compromising lifestyle and then reinforces that pattern through various channels of social influence. In a study of 293 inner-city injecting drug users (IDUs), Latkin and colleagues (333) found that network density and size of drug-using subnetworks were positively associated with frequency of drug use. That same study found that injectors whose personal networks contained a noninjecting spouse/lover/sex partner injected less frequently, suggesting that some kinds of ties may be protective against high-risk exposure. These findings are corroborated by studies that show that supportive ties ameliorate the influence of high-risk environments on drug use (334, 335).

Numerous commentators have noted that one of the reasons that behavior change interventions appear to be ineffective among injection drug users is that the social networks of IDUs themselves are stark barriers to that effectiveness. This has led to calls for interventions that attempt to work with, rather than around, those networks (336, 337). An innovative example comes from Kelly and collaborators (338), who used key opinion leaders in high-risk social settings to attempt to change norms around high-risk behaviors.

The study of the role of social support and infectious disease has also led to important evidence regarding potential biophysiological pathways through which social support may operate. For example, in one of the only studies of its type, Theorell and colleagues (339) followed HIV-infected men in Sweden and tracked their decline in CD4 count over a 5-year period. This group found that men who reported lower "availability of attachments" at baseline declined more rapidly in CD4 levels, indicating the possibility that social support mediates primary immune system parameters.

Social scientists such as Roderick Wallace have noted that at the intersection of social networks and physical locations in space, sociogeographic matrices are formed through which and across which epidemics such as HIV travel (340). The structure of overlapping networks located in space acts as a system of conduits allowing for epidemic spread. At the micro level this phenomenon is visible in the rapid rates of saturation that occur within subnetworks of high-risk individuals. This phenomenon is also visible at a more macro level. For example, Hunt et al. (341) demonstrate how patterns of HIV transmission vary within regions of Uganda according to patterns of migrant labor use; these labor market patterns form the social network infrastructure, which creates opportunities for the spread of the epidemic.

SOCIAL SUPPORT AND OTHER INFECTIOUS DISEASES

That social contacts may confer a generalized host resistance against the development of infectious disease was suggested in early papers by, among others, Cassel (4) and Selye (342). More recently, a solid foundation of evidence has begun to show that social support appears to alter primary immune system parameters that regulate host resistance (27, 168, 343, 344). Cohen and colleagues (120) conducted an experiment to test the hypothesis that the diversity of network ties is related to susceptibility to the common cold. In this experiment, after reporting the extent of participation in 12 types of social ties (e.g., spouse, parent, friend, workmate, member of social group), subjects were given nasal drops containing one of two rhinoviruses, and were monitored for the development of a common cold. Results indicated that those with more types of social ties were less susceptible to common colds, produced less mucus, fought infection more efficiently, and shed less virus even after controlling for prechallenge virus-specific antibody, virus type, age, sex, season, body mass index, education, and race. Susceptibility to colds decreased in a dose-response manner with increased diversity of the social network.

If indeed social integration and participation are associated with changes in immune system functioning, the implications are far reaching. First, immune system functioning is directly linked to the development of infectious diseases, allergies, autoimmune diseases, and cancer (24). While the evidence is less compelling that lack of social support increases risk of cancer and autoimmune disease, the evidence compiled by Cohen and others has important implications for the likelihood of this effect. In addition, the discovery of the influence of social support on neuroendocrine function suggests that social support processes may contribute to the pathogenesis of cardiovascular disease due to the influence of immune system function on the health of the arterial system and on hemodynamic processes.

CONCLUSIONS

In this chapter, we have reframed discussions of the impact that different qualities or dimensions of social relationships have on health by placing them in the larger context of social networks. This is a very large undertaking, and when coupled with a search for a deeper understanding of how social networks influence health it becomes even larger. It is incumbent on the investigators wishing to work in this area to recast general ideas about networks and health into specific hypotheses, which will be testable through the explicit identification and articulation of theories, pathways, and mechanisms through which social networks impact health. It is unlikely that any single measure or study design will be useful for all purposes, diseases, or behaviors. Rather, by articulating the "upstream" contextual influences of network structure and "downstream" pathways by which networks more directly and proximately influence health, investigators will make significant progress. Much of this work invoked earlier theories and often used measures that conceptually blurred domains of networks and functional aspects of such ties.

The greatest challenges facing us in network epidemiology, however, do not have to do with measuring the exposure. The last decade's work made this problem much more tractable. Rather, the problems that confront the field now have to do with reconciling with the observational evidence the null or modest effects of interventions based on rigorous randomized trials. One clear interpretation of these differences in associations by study design is that there is not a causal association between networks and changes in health or health behaviors. Another interpretation is that we have not succeeded in changing network structure or support enough or at the correct etiologic period to impact the health outcomes we are hoping to influence. Furthermore, arguments about the causal effects across multiple degrees of separation in observational studies from the Framingham cohorts have suggested that while the findings may not be as robust to confounding as investigators have reported, even conservative estimates support arguments that some network effects are likely to be causing behavioral change (3). The jury is out. Developing interventions that can change networks and support during an important etiologic period, and conducting observational studies that deal effectively with confounding and reverse causation remain the most important challenges for future investigators.

REFERENCES

1. Berkman LF. Social epidemiology: Social determinants of health in the United States: Are we losing ground? Annu Rev Public Health. 2009;30(1):27–41.
2. VanderWeele TJ, Hawkley LC, Thisted RA, Cacioppo JT. A marginal structural model analysis for loneliness: implications for intervention trials and clinical practice. J Consult Clin Psychol. 2011;79(2):225–35.
3. VanderWeele TJ. Inference for influence over multiple degrees of separation on a social network. Stat Med. 2013;32(4):591–6.
4. Cassel J. The contribution of the social environment to host resistance. Am J Epidemiol. 1976;104(2):107–23.
5. Cobb S. Social support as a moderator of life stress. Psychosom Med. 1976;38(5):300–14.
6. Berkman LF, Syme SL. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. Am J Epidemiol. 1979;109(2):186–204.

7. Blazer DG. Social support and mortality in an elderly community population. Am J Epidemiol. 1982;115(5):684–94.
8. House JS, Robbins C, Metzner HL. The association of social relationships and activities with mortality: prospective evidence from the Tecumseh Community Health Study. Am J Epidemiol. 1982;116(1):123–40.
9. Holt-Lunstad J, Smith TB, Layton JB. Social relationships and mortality risk: a meta-analytic review. PLoS Med. 2010;7(7):e1000316.
10. Berkman LF. The role of social relations in health promotion. Psychosom Med. 1995;57(3):245–54.
11. Berkman LF, Glass T, Brissette I, Seeman TE. From social integration to health: Durkheim in the new millennium. Soc Sci Med. 2000;51(6):843–57.
12. Cacioppo S, Cacioppo JT. Decoding the invisible forces of social connections. Front Integr Neurosci. 2012;6:51.
13. Cacioppo JT, Hawkley LC. Perceived social isolation and cognition. Trends Cogn Sci. 2009;13(10):447–54.
14. Cacioppo JT, Fowler JH, Christakis NA. Alone in the crowd: The structure and spread of loneliness in a large social network. J Pers Soc Psychol. 2009;97(6):977–91.
15. Cacioppo JT, Decety J. Social neuroscience: challenges and opportunities in the study of complex behavior. Ann N Y Acad Sci. 2011;1224:162–73.
16. Cacioppo JT, Reis HT, Zautra AJ. Social resilience: the value of social fitness with an application to the military. Am Psychol. 2011;66(1):43–51.
17. Hawkley LC, Cacioppo JT. Loneliness matters: a theoretical and empirical review of consequences and mechanisms. Ann Behav Med. 2010;40(2):218–27.
18. Cacioppo JT, Hawkley LC, Norman GJ, Berntson GG. Social isolation. Ann N Y Acad Sci. 2011;1231:17–22.
19. Fowler JH, Christakis NA. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. BMJ. BMJ Group; 2008;337:a2338–8.
20. Fowler JH, Christakis NA. Estimating peer effects on health in social networks: A response to Cohen-Cole and Fletcher; and Trogdon, Nonnemaker, and Pais. J Health Econ. 2008;27(5):1400–5.
21. Christakis NA. Social networks and collateral health effects. BMJ. 2004;329(7459):184–5.
22. Christakis NA, Fowler JH. Social contagion theory: examining dynamic social networks and human behavior. Stat Med. 2013;32(4):556–77.
23. Smith KP, Christakis NA. Social networks and health. Annu Rev Sociol. 2008;34(1):405–29.
24. Cohen S. Psychosocial models of the role of social support in the etiology of physical disease. Health Psychol. 1988;7(3):269–97.
25. Cohen S, Kaplan JR, Cunnick JE, Manuck SB, Rabin BS. Chronic social stress, affiliation, and cellular immune response in nonhuman primates. Psychol Sci. 1992;3(5):301–4.
26. Cacioppo JT. Social neuroscience: autonomic, neuroendocrine, and immune responses to stress. Psychophysiology. 1994;31(2):113–28.
27. Kiecolt-Glaser JK, Malarkey W, Cacioppo JT, Glaser R. Stressful personal relationships: immune and endocrine function. In: Glaser R, Kiecolt-Glaser JK, editors. Handbook of human stress and immunity. San Diego, CA: Academic Press; 2001. pp. 321–39.
28. Sgoutas-Emch SA, Cacioppo JT, Uchino BN, Malarkey W, Pearl D, Kiecolt-Glaser JK, et al. The effects of an acute psychological stressor on cardiovascular, endocrine, and cellular immune response: A prospective study of individuals high and low in heart rate reactivity. Psychophysiology. 1994;31(3):264–71.
29. Uchino BN, Kiecolt-Glaser JK, Cacioppo JT. Age-related changes in cardiovascular response as a function of a chronic stressor and social support. J Pers Soc Psychol. 1992;63:839–9.
30. McEwen BS. Allostasis and allostatic load: implications for neuropsychopharmacology. Neuropsychopharmacol. 2000;22(2):108–24.

31. McEwen BS. Stress, adaptation, and disease: allostasis and allostatic load. Ann N Y Acad Sci. 1998;840(1): 33–44.
32. McEwen BS, Seeman T. Protective and damaging effects of mediators of stress: elaborating and testing the concepts of allostasis and allostatic load. Ann N Y Acad Sci. 1999; 896(1):30–47.
33. Barnes JA. Class and committees in a Norwegian island parish. Hum Relat. 1954;7(1):39–58.
34. Bott E. Family and social network. London: Tavistock Press; 1957.
35. Wellman B. An egocentric network tale: comment on Bien et al. (1991). Soc Networks. 1993;15:423–36.
36. Laumann EO. Bonds of pluralism. New York: John Wiley; 1973.
37. Wellman B, Leighton B. Networks, neighborhoods, and communities: Approaches to the study of the community question. Urban Aff Rev. 1979;14(3):363–90.
38. Granovetter MS. The strength of weak ties. Am J Sociol. 1973;78:1360–80.
39. Fischer CS, Stueve C, Jones LM, Jackson RM. Networks and places: Social relations in the urban setting. New York: Free Press; 1977.
40. Fischer CS. To dwell among friends: Personal networks in town and city. Chicago: University of Chicago Press; 1982.
41. Hall B, Wellman B. Social networks and social support. In: Cohen S, Syme SL, editors. Social support and Health. Orlando: Academic Press; 1985. pp. 23–41.
42. Wellman B, Carrington PJ. Networks as personal communities. In: Wellman B, Berkowitz SD, editors. Social structures: a network approach: structural analysis in the social sciences. New York: Cambridge University Press; 1988. pp. 130–84.
43. Durkheim E. The division of labor in society. New York: Free Press; 1895.
44. Durkheim E. Suicide: a study in sociology. Glencoe, IL: Free Press; 1897.
45. Link BG, Phelan J. Social conditions as fundamental causes of disease. J Health Soc Behav. 1995;Spec No:80–94.
46. Durkheim E. The rules of sociological method. Lukes S, editor. New York: Free Press; 1895.
47. Bierstedt R. Émile Durkheim. London: Weidenfeld and Nicolson; 1966. p. 1.
48. Turner JH, Beeghley L, Powers CH. The emergence of sociological theory. Chicago, IL: Dorsey Press; 1989.
49. LaCapra D. Emile Durkheim: sociologist and philosopher. Cornell University Press. Ithaca, NY; 1972.
50. Bowlby J. A secure base. London: Routledge; 1988.
51. Storr A. John Bowlby, Munks Roll. London: Royal College of Physicians; 1991.
52. Fonagy P. Patterns of attachment, interpersonal relationships and health. In: Blane D, Brunner E, Wilkinson R, editors. Health and social organization: towards health policy for the twenty-first century. London: Routledge Press; 1996. pp. 125–51.
53. Bowlby J. Attachment. London: Hogarth Press; 1969.
54. Bowlby J. Separation—anxiety and anger. London: Hogarth Press; 1973.
55. Bowlby J. Loss—sadness and depression. London: Hogarth Press; 1980.
56. Kubzansky LD, Mendes WB, Appleton AA, Block J, Adler GK. A heartfelt response: oxytocin effects on response to social stress in men and women. Biol Psychol. 2012;90(1):1–9.
57. Appleton AA, Buka SL, Loucks EB, Rimm EB, Martin LT, Kubzansky LD. A prospective study of positive early-life psychosocial factors and favorable cardiovascular risk in adulthood. Circulation. 2013;127(8):905–12.
58. Appleton AA, Buka SL, McCormick MC, Koenen KC, Loucks EB, Kubzansky LD. The association between childhood emotional functioning and adulthood inflammation is modified by early-life socioeconomic status. Health Psychol. 2012;31(4):413–22.

59. Maselko J, Kubzansky L, Lipsitt L, Buka SL. Mother's affection at 8 months predicts emotional distress in adulthood. J Epidemiol Community Health. 2011;65(7):621–5.
60. Repetti RL, Taylor SE, Seeman TE. Risky families: family social environments and the mental and physical health of offspring. Psychol Bull. 2002;128(2):330.
61. Taylor SE, Klein LC, Lewis BP, Gruenewald TL, Gurung RA, Updegraff JA. Biobehavioral responses to stress in females: tend-and-befriend, not fight-or-flight. Psychol Rev. 2000;107(3):411–29.
62. Drury SS, Theall K, Gleason MM, Smyke AT, De Vivo I, Wong JYY, et al. Telomere length and early severe social deprivation: linking early adversity and cellular aging. Mol Psychiatry. 2011;17(7):719–27.
63. Fox SE, Levitt P, Nelson CA III. How the timing and quality of early experiences influence the development of brain architecture. Child Dev. 2010;81(1):28–40.
64. Nelson CA, Zeanah CH, Fox NA, Marshall PJ, Smyke AT, Guthrie D. Cognitive recovery in socially deprived young children: the Bucharest early intervention project. Science. 2007;318(5858):1937–40.
65. Holmes J. John Bowlby and attachment theory. London: Routledge; 1993.
66. Cacioppo JT, Hughes ME, Waite LJ, Hawkley LC, Thisted RA. Loneliness as a specific risk factor for depressive symptoms: cross-sectional and longitudinal analyses. Psychol Aging. 2006;21(1):140–51.
67. Cacioppo JT, Cacioppo S. The phenotype of loneliness. Eur J Dev Psychol. 2012;9(4):446–52.
68. Kuh D, Ben-Shlomo Y. A life course approach to chronic disease epidemiology. 2nd ed. New York: Oxford University Press; 2005.
69. Kuh D, Ben-Shlomo Y, Lynch J, Hallqvist J, Power C. Life course epidemiology. J Epidemiol Community Health. 2003;57(10):778–83.
70. Granovetter M. The strength of weak ties: a network theory revisited. In: Marsden P, Lin N, editors. Social structure and network analysis. Beverly Hills, CA: Sage; 1982. pp. 105–30.
71. Marsden PV. Network studies of social influence. In: Wasserman S, Galaskiewicz J, editors. Advances in social network analysis: research in the social and behavioral sciences. Thousand Oaks, CA: Sage; 1994. pp. 3–25.
72. Rosenquist JN, Murabito J, Fowler JH, Christakis NA. The spread of alcohol consumption in a large social network. Ann Intern Med. 2010;152(7):426.
73. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. N Engl J Med. 2008;358(21):2249–58.
74. Valente TW, Watkins SC, Jato MN. Social network associations with contraceptive use among Cameroonian women in voluntary associations. Soc Sci Med. 1997;45(5):677–87.
75. Helleringer S, Kohler H-P, Chimbiri A, Chatonda P, Mkandawire J. The Likoma Network Study: context, data collection, and initial results. Demographic Res. 2009;21:427–68.
76. Morris M, Kurth AE, Hamilton DT, Moody J, Wakefield S. Concurrent partnerships and HIV prevalence disparities by race: linking science and public health practice. Am J Public Health. 2009;99(6):1023–31.
77. Morris M, Podhista C, Wawer MJ, Handcock MS. Bridge populations in the spread of HIV/AIDS in Thailand. AIDS. 1996;10(11):1265–71.
78. Kohler H-P, Behrman JR, Watkins SC. Social networks and HIV/AIDs risk perceptions. Demography. 2007;44(1):1–33.
79. Helleringer S, Kohler H-P. Sexual network structure and the spread of HIV in Africa: evidence from Likoma Island, Malawi. AIDS. 2007;21(17):2323–32.
80. Christakis NA. Putting the social into science: forget about nature vs. nurture; the answer lies in between. Time. 2011 December 19:28.
81. Bearman PS, Moody J, Stovel K. Chains of affection: the structure of adolescent romantic and sexual networks. Am J Sociol. 2004;110(1):44–91.
82. Bearman P, Moody J, Faris R. Networks and history. Complexity. 2003;8(1):61–71.

83. Bearman PS, Moody J. Suicide and friendships among American adolescents. Am J Public Health. 2004;94(1):89–95.
84. Moody J, Feinberg ME, Osgood DW, Gest SD. Mining the network: peers and adolescent health. J Adolesc Health. 2010;47(4):324–6.
85. Moody J. Peer influence groups: identifying dense clusters in large networks. Soc Networks. 2001;23(4):261–83.
86. Moody J, Brynildsen WD, Osgood DW, Feinberg ME, Gest S. Popularity trajectories and substance use in early adolescence. Soc Networks. 2011;33(2):101–12.
87. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med. 2007;357(4):370–9.
88. Valente TW, Hoffman BR, Ritt-Olson A, Lichstein K, Anderson Johnson C. Effects of a social-network method for group assignment strategies on peer-led tobacco prevention programs in schools. Am J Public Health. 2003;93(11):1837.
89. Valente TW. Social networks and health: models, methods, and applications. New York: Oxford University Press; 2010.
90. Valente TW, Ritt-Olson A, Stacy A, Unger JB, Okamoto J, Sussman S. Peer acceleration: effects of a social network tailored substance abuse prevention program among high-risk adolescents. Addiction. 2007;102(11):1804–15.
91. Valente TW, Saba WP. Mass media and interpersonal influence in a reproductive health communication campaign in Bolivia. Comm Res. 1998;25(1):96–124.
92. Valente TW, Fujimoto K, Soto D, Ritt-Olson A, Unger JB. A comparison of peer influence measures as predictors of smoking among predominately Hispanic/Latino high school adolescents. J Adolesc Health. 2013;52(3):358–64.
93. Valente TW, Fujimoto K. Bridging: locating critical connectors in a network. Soc Networks. 2010;32(3):212–20.
94. Valente TW. Network interventions. Science. 2012;337(6090):49–53.
95. Valente TW. "Network models and methods for studying the diffusion of innovations." In Carrington PJ, Scott JJ, Wasserman S, editors. Models and methods in social network analysis. New York: Cambridge University Press; 2005. pp. 98–116.
96. Goodreau SM, Kitts JA, Morris M. Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks. Demography. 2009;46(1):103–25.
97. Helleringer S, Kohler H-P. Cross-sectional research design and relatively low HIV incidence, rather than blood exposures, explain the peripheral location of HIV cases within the sexual networks observed on Likoma. AIDS. 2008;22(11):1378–9.
98. Kohler HP, Behrman JR. Empirical assessments of social networks, fertility and family planning programs: nonlinearities and their implications. Demographic Res. 2000;3(7):79–126.
99. Kohler HP. Learning in social networks and contraceptive choice. Demography. 1997;34(3):369–83.
100. Helleringer S, Kohler H-P. Social networks, perceptions of risk, and changing attitudes towards HIV/AIDS: new evidence from a longitudinal study using fixed-effects analysis. Population Studies. 2005;59(3):265–82.
101. Kohler HP, Behrman JR, Watkins SC. The density of social networks and fertility decisions: evidence from South Nyanza District, Kenya. Demography. 2001.
102. Kohler HP, Helleringer S, Behrman JR, Watkins SC. The social and the sexual: networks in contemporary demographic research. PSC Working Paper Series. 2013.
103. House JS, Landis KR, Umberson D. Social relationships and health. Science. 1988;241(4865):540–5.
104. Berkman LF. Social networks, support, and health: taking the next step forward. Am J Epidemiol. 1986;123(4):559–62.

105. Antonucci TC, Jackson JS. The role of reciprocity in social support. In: Sarason BR, Sarason IG, Pierce GR, editors. Social support: an interactional view. New York: John Wiley & Sons; 1990. pp. 173–98.

106. Antonucci TC. Measure social support networks: hierarchical mapping techniques. Generations. 1986;10:10–2.

107. Antonucci TC, Akiyama H. Social networks in adult life and a preliminary examination of the convoy model. J Gerontol. 1987;42(5):519–27.

108. Kahn RL. Aging and social support. In: Riley MW, editor. Aging from birth to death: an interdisciplinary perspective. Boulder, CO: Westview; 1979. pp. 72–92.

109. Lin N, Dean A. Social support and depression: a panel study. Soc Psychiatry. 1984;19(2):83–91.

110. Dean A, Lin N. The stress-buffering role of social support: problems and prospects for systematic investigation. J Nerv Ment Dis. 1977;165(6):403.

111. Lin N, Woelfel MW, Light SC. The buffering effect of social support subsequent to an important life event. J Health Soc Behav. 1985;26(3):247–63.

112. Lin N, Dean A, Ensel WM. Social support, life events, and depression. New York: Academic Press; 1986.

113. Lin N, Dean A, Ensel WM. Social support scales: a methodological note. Schizophr Bull. 1981;7(1):73–89.

114. House JS. Work stress and social support. Reading, MA: Addison Wesley Publishing Company; 1981.

115. House JS, Kahn R. Measures and concepts of social support. In: Cohen S, Syme SL, editors. Social support and health. Orlando: Academic Press; 1985.

116. LaRocca JM, House JS, French JRJ. Social support, occupational stress and health. J Health Soc Behav. 1980;21:202–8.

117. Cohen S, Janicki-Deverts D, Miller GE. Psychological stress and disease. JAMA. 2007;298(14):1685–7.

118. Cohen S. Social relationships and health. Am Psychol. 2004;59(8):676–84.

119. Cohen S, Janicki-Deverts D. Can we improve our physical health by altering our social networks? Perspect Psychol Sci. 2009;4(4):375–8.

120. Cohen S, Doyle WJ, Skoner DP, Rabin BS, Gwaltney JM. Social ties and susceptibility to the common cold. JAMA. 1997;277(24):1940–4.

121. Cohen S, Lemay EP. Why would social networks be linked to affect and health practices? Health Psychol. 2007;26(4):410–7.

122. Rook KS. Social relationships as a source of companionship: implications for older adults' psychological well being. In: Sarason BR, Sarason IG, Pierce GR, editors. Social support: an interactional view. New York: John Wiley & Sons; 1990. pp. 221–50.

123. Rook KS. Social support versus companionship: effects on life stress, loneliness, and evaluations by others. J Pers Soc Psychol.1987;52(6):1132–47.

124. Rook KS. The negative side of social interaction: impact on psychological well-being. J Pers Soc Psychol. 1984;46(5):1097.

125. Sarason BR, Sarason IG, Pierce GR. Social support: an interactional view. New York: John Wiley & Sons; 1990.

126. Sarason IG, Levine HM, Basham RB. Assessing social support: the social support questionnaire. J Pers Soc Psychol. 1983;44(1):127–39.

127. Schaefer C, Coyne JC, Lazarus RS. The health-related functions of social support. J Behav Med. 1981;4(4):381–406.

128. Kahn RL, Antonucci TC. Convoys over the lifecourse: attachment, roles and social support. In: Baltes PB, Brim O, editors. Life span development and behavior. New York: Academic Press; 1980. pp. 253–86.

129. Watts D. Six degrees: the science of a connected age. New York: W. W. Norton & Company; 2004.

130. Fowler JH, Settle JE, Christakis NA. Correlated genotypes in friendship networks. PNAS. 2011;108(5):1993–7.

131. Cassels S, Clark SJ, Morris M. Mathematical models for HIV transmission dynamics: tools for social and behavioral science research. J Acquir Immune Defic Syndr. 2008;47(Suppl 1):S34–9.

132. Luxton M. More than a labor of love. Toronto: Women's Press; 1980.

133. Krause N, Markides K. Measuring social support among older adults. Int J Aging Hum Dev. 1995;30(1): 37–53.

134. Bodemann YM. Relations of product and class rule: the basis of patron/clientage. In: Wellman B, Berkowitz SD, editors. Social structures: a network approach. Cambridge, UK: Cambridge University Press; 1988. pp. 198–220.

135. Belle DE. The impact of poverty on social networks and supports. Marriage Fam Rev. 1983;5(4):89–103.

136. Fischer CS. The 2004 GSS finding of shrunken social networks: An artifact? Am Sociol Rev. 2009;74(4): 657–69.

137. Putnam RD. Bowling alone: The collapse and revival of American community. New York: Simon & Schuster; 2001.

138. Cohen S, Underwood L, Gottlieb B, editors. Social support measurement and interventions: a guide for health and social scientists. New York: Oxford; 2000.

139. Cohen S, Syme SL, editors. Social support and health. San Diego, CA: Academic Press; 1985.

140. Thoits PA. Stress, coping, and social support processes: where are we? What next? J Health Soc Behav. 1995;35(Spec No):53–79.

141. George LK. Caregiver burden: conflict between norms of reciprocity and solidarity. In: Pillemar KA, Wolf RD, editors. Conflict and abuse in families of the elderly: theory, research and intervention. Boston: Auburn House; 1986. pp. 67–92.

142. Dunkel-Schetter C, Bennett TL. Differentiating the cognitive and behavioral aspects of social support. In: Sarason BR, Sarason IG, Pearson GR, editors. Social support: an interactional view. New York: John Wiley & Sons; 1990. pp. 267–96.

143. Erickson BH. The relational basis of attitudes. In: Wellman B, Berkowitz SD, editors. Social structures: a network approach. New York: Cambridge University Press; 1988. pp. 99–121.

144. Glymour MM, Weuve J, Fay ME, Glass T, Berkman LF. Social ties and cognitive recovery after stroke: does social integration promote cognitive resilience? Neuroepidemiology. 2008;31(1):10–20.

145. Glymour MM, Maselko J, Gilman SE, Patton KK, Avendano M. Depressive symptoms predict incident stroke independently of memory impairments. Neurology. 2010;75(23):2063–70.

146. Tun PA, Miller-Martinez D, Lachman ME, Seeman T. Social strain and executive function across the lifespan: the dark (and light) sides of social engagement. Aging Neuropsychol C. 2013;20(3): 320–38.

147. Morris M. Data driven network models for the spread of infectious disease. In: Mollison D, editor. Epidemic models: their structure and relation to data. Cambridge, UK: Cambridge University Press; 1995. pp. 302–22.

148. Morris M. Epidemiology and social networks: modeling structured diffusion. In: Wasserman S, Galaskiewicz J, editors. Advances in social network analysis: research in the social and behavioral sciences. Thousand Oaks, CA: Sage; 1994. pp. 26–52.

149. Morris M. Network epidemiology: a handbook for survey design and data collection. New York: Oxford University Press; 2004.

150. Laumann E, Gagnon J, Michaels S, Michael R, Coleman J. Monitoring the AIDS epidemic in the United States: a network approach. Science. 1989;244(4909):1186–9.

151. Friedman SR. Promising social network research results and suggestions for a research agenda. NIDA research monograph. 1995.

152. Klovdahl AS. Social networks and the spread of infectious diseases: the AIDS example. Soc Sci Med. 1985;21(11):1203–16.

153. Moody J, McFarland D, Bender deMoll S. Dynamic network visualization. Am J Sociol. 2005;110(4):1206–41.
154. Behrman JR, Kohler HP, Watkins SC. Lessons from empirical network analyses on matters of life and death in East Africa. California Center for Population Research Online Working Paper Series 2008.
155. Seeman TE, Berkman LF. Structural characteristics of social networks and their relationship with social support in the elderly: who provides support. Soc Sci Med. Elsevier; 1988;26(7):737–49.
156. Seeman TE, Berkman LF, Kohout F, Lacroix A, Glynn R, Blazer D. Intercommunity variations in the association between social ties and mortality in the elderly. Ann Epidemiol. 1993;3(4):325–35.
157. Norman GJ, Hawkley L, Ball A, Berntson GG, Cacioppo JT. Perceived social isolation moderates the relationship between early childhood trauma and pulse pressure in older adults. Int J Psychophysiol. 2013.
158. Krause N. Longitudinal study of social support and meaning in life. Psychol Aging. 2007;22(3):456–69.
159. Cho HJ, Bower JE, Kiefe CI, Seeman TE, Irwin MR. Early life stress and inflammatory mechanisms of fatigue in the Coronary Artery Risk Development in Young Adults (CARDIA) study. Brain Behav Immun. 2012;26(6):859–65.
160. Taylor D, Bury M. Chronic illness, expert patients and care transition. Sociol Health Ill. 2007;29(1):27–45.
161. Seeman T, Dostálek L, Gilík J. Control of hypertension in treated children and its association with target organ damage. Am J Hypertens. 2012;25(3):389–95.
162. Seeman T, Gilík J. Long-term control of ambulatory hypertension in children: improving with time but still not achieving new blood pressure goals. Am J Hypertens. 2013;26(7):939–45.
163. Chiang JJ, Eisenberger NI, Seeman TE, Taylor SE. Negative and competitive social interactions are related to heightened proinflammatory cytokine activity. PNAS. 2012;109(6):1878–82.
164. Friedman EM, Karlamangla AS, Almeida DM, Seeman TE. Social strain and cortisol regulation in midlife in the US. Soc Sci Med. 2012;74(4):607–15.
165. Birditt KS, Antonucci TC, Tighe L. Enacted support during stressful life events in middle and older adulthood: An examination of the interpersonal context. Psychol Aging. 2012;27(3):728–41.
166. Birditt KS, Jackey LMH, Antonucci TC. Longitudinal patterns of negative relationship quality across adulthood. J Gerontol B Psychol Sci Soc Sci. 2009;64B(1):55–64.
167. Ajrouch KJ, Abdulrahim S, Antonucci TC. Stress, social relations, and psychological health over the life course. GeroPsych. 2013;26(1):15–27.
168. Uchino BN, Cacioppo JT, Kiecolt-Glaser JK. The relationship between social support and physiological processes: A review with emphasis on underlying mechanisms and implications for health. Psychol Bull. 1996;119(3):488–531.
169. Knox SS, Uvnäs-Moberg K. Social isolation and cardiovascular disease: an atherosclerotic pathway? Psychoneuroendocrinology. 1998;23(8):877–90.
170. Hawkley LC, Thisted RA, Masi CM, Cacioppo JT. Loneliness predicts increased blood pressure: 5-year cross-lagged analyses in middle-aged and older adults. Psychol Aging. 2010;25(1):132–41.
171. Hawkley LC, Browne MW, Cacioppo JT. How can I connect with thee? Let me count the ways. Psychol Sci. 2005;16(10):798–804.
172. Hawkley LC, Preacher KJ, Cacioppo JT. Loneliness impairs daytime functioning but not sleep duration. Health Psychol. 2010;29(2):124–9.
173. Fujimoto K, Unger JB, Valente TW. A network method of measuring affiliation-based peer influence: assessing the influences of teammates' smoking on adolescent smoking. Child Dev. 2012;83(2):442–51.
174. Alexander C, Piazza M, Mekos D, Valente T. Peers, schools, and adolescent cigarette smoking. J Adolesc Health. 2001;29(1):22–30.
175. Fujimoto K, Valente TW. Social network influences on adolescent substance use: disentangling structural equivalence from cohesion. Soc Sci Med. 2012;74(12):1952–60.

176. Lakon CM, Valente TW. Social integration in friendship networks: the synergy of network structure and peer influence in relation to cigarette smoking among high risk adolescents. Soc Sci Med. 2012;74(9):1407–17.

177. Valente TW, Fosados R. Diffusion of innovations and network segmentation: The part played by people in promoting health. Sex Transm Dis. 2006; 33(Suppl):S23–S31.

178. Newsom JT, Rook KS, Nishishiba M, Sorkin DH, Mahan TL. Understanding the relative importance of positive and negative social exchanges: Examining specific domains and appraisals. J Gerontol. 2005;60B(6):P304–12.

179. August KJ, Rook KS, Newsom JT. The joint effects of life stress and negative social exchanges on emotional distress. J Gerontol. 2007;62B(5):S304–14.

180. Chronis AM, Lahey BB, Pelham WEJ, Williams SH, Baumann BL, Kipp H, et al. Maternal depression and early positive parenting predict future conduct problems in young children with attention-deficit/hyperactivity disorder. Dev Psychol. 2007;43(1):70–82.

181. Plaisier I, de Bruijn JG, de Graaf R, Have ten M, Beekman AT, Penninx BW. The contribution of working conditions and social support to the onset of depressive and anxiety disorders among male and female employees. Soc Sci Med. 2007;64(2):401–10.

182. Desteno D, Gross JJ, Kubzansky L. Affective science and health: the importance of emotion and emotion regulation. Health Psychol. 2013;32(5):474–86.

183. Frederickson BL. The broaden-and-build theory of positive emotions. In: Huppert FA, Baylis N, B K, editors. The science of well-being. New York: Oxford University Press; 2005. pp. 217–38.

184. Grembowski D, Patrick D, Diehr P, Durham M, Beresford S, Kay E, et al. Self-efficacy and health behavior among older adults. J Health Soc Behav. 1993;34(2):89–104.

185. McAuley E. Self-efficacy, physical activity, and aging. In: Kelly JR, editor. Activity and aging: staying involved in later life. Newbury Park, CA: Sage; 1993. pp. 187–206.

186. Mendes de Leon CF, Seeman TE, Baker DI, Richardson ED, Tinetti ME. Self-efficacy, physical decline, and change in functioning in community-living elders: A prospective study. J Gerontol B Psychol Sci Soc Sci. 1996;51B(4):S183–90.

187. Seeman TE, Rodin J, Albert MA. Self-efficacy and functional ability: how beliefs relate to cognitive and physical performance. J Aging Health. 1993;5:455–74.

188. Tinetti ME, Powell L. Fear of falling and low self-efficacy: a cause of dependence in elderly persons. J Gerontol. 1993;48(Special):35–8.

189. Cutrona CE, Troutman BR. Social support, infant temperament, and parenting self-efficacy: a mediational model of postpartum depression. Child Dev. 1986;57(6):1507–18.

190. Major B, Cozzarelli C, Sciacchitano AM, Cooper ML, Testa M, Mueller PM. Perceived social support, self-efficacy, and adjustment to abortion. J Pers Soc Psychol. 1990;59(3):452–63.

191. Gulliver SB, Hughes JR, Solomon LJ, Dey AN. An investigation of self-efficacy, partner support and daily stresses as predictors of relapse to smoking in self-quitters. Addiction. 1995;90(6):767–72.

192. McFarlane AH, Bellissimo A, Norman GR. The role of family and peers in social self-efficacy: links to depression in adolescence. Am J Orthopsychiatry. 1995;65(3):402–10.

193. Duncan TE, McAuley E. Social support and efficacy cognitions in exercise adherence: a latent growth curve analysis. J Behav Med. 1993;16(2):199–218.

194. McAvay GJ, Seeman TE, Rodin J. A longitudinal study of change in domain-specific self-efficacy among older adults. J Gerontol B Psychol Sci Soc Sci. 1996;51B(5):P243–53.

195. Holahan CJ, Moos RH. Personal and contextual determinants of coping strategies. J Pers Soc Psychol. 1987;52(5):946–55.

196. Wolf TM, Balson PM, Morse EV, Simon PM, Gaumer RH, Dralle PW, et al. Relationship of coping style to affective state and perceived social support in asymptomatic and symptomatic HIV-infected persons: implications for clinical management. J Clin Psychiat. 1991;52(4):171–3.

197. Dunkel-Schetter C, Folkman S, Lazarus RS. Correlates of social support receipt. J Pers Soc Psychol. 1987;53(1):71–80.

198. Dunkel-Schetter C, Feinstein LG, Taylor SE, Falke RL. Patterns of coping with cancer. Health Psychol. 1992;11:79–87.

199. Bowling A, Browne PD. Social networks, health, and emotional well-being among the oldest old in London. J Gerontol. 1991;46(1):S20–S32.

200. Holahan CJ, Moos RH, Holahan CK, Brennan PL. Social support, coping, and depressive symptoms in a late-middle-aged sample of patients reporting cardiac illness. Health Psychol. 1995;14(2):152–63.

201. Lomauro TA. Social support, health locus-of-control, and coping style and their relationship to depression among stroke victims. DAI. 1990;51:2628.

202. Matt GE, Dean A. Social support from friends and psychological distress among elderly persons: moderator effects of age. J Health Soc Behav. 1993;34(3):187–200.

203. Morris PL, Robinson RG, Raphael B, Bishop D. The relationship between the perception of social support and post-stroke depression in hospitalized patients. Psychiatry. 1991;54(3):306–16.

204. George LK, Blazer DG, Hughes DC, Fowler N. Social support and the outcome of major depression. Br J Psychiatry. 1989;154:478–85.

205. Turner RJ. Direct, indirect and moderating effects of social support upon psychological distress and associated condition. In: Kaplan H, editor. Psychosocial stress: trends in theory and research. New York: Academic Press; 1983. pp. 105–55.

206. Oxman TE, Berkman LF, Kasl S, Freeman DH, Barrett J. Social support and depressive symptoms in the elderly. Am J Epidemiol. 1992;135(4):356–68.

207. Blazer DG. Depression and social support in late life: a clear but not obvious relationship. Aging Ment Health. 2005;9(6):497–9.

208. Blazer DG, Hybels CF. Origins of depression in later life. Psychol Med. 2005;35(9):1241–52.

209. Lett HS, Blumenthal JA, Babyak MA, Catellier DJ, Carney RM, Berkman LF, et al. Social support and prognosis in patients at increased psychosocial risk recovering from myocardial infarction. Health Psychol. 2007;26(4):418–27.

210. Paykel ES. Life events, social support and depression. Acta Psychiatr Scand. 1994;89(s377):50–8.

211. Vilhjalmsson R. Life stress, social support and clinical depression: a reanalysis of the literature. Soc Sci Med. 1993;37(3):331–42.

212. Murphy E. Social origins of depression in old age. Br J Psychiatry. 1982;141(2):135–42.

213. Henderson S. Social relationships, adversity and neurosis: an analysis of prospective observations. Br J Psychiatry. 1981;138:391–8.

214. Loucks EB, Buka SL, Rogers ML, Liu T, Kawachi I, Kubzansky LD, et al. Education and coronary heart disease risk associations may be affected by early-life common prior causes: a propensity matching analysis. Ann Epidemiol. 2012;22(4):221–32.

215. Loucks EB, Berkman LF, Gruenewald TL, Seeman TE. Relation of social integration to inflammatory marker concentrations in men and women 70 to 79 years. Am J Cardiol. 2006;97(7):1010–6.

216. Loucks EB, Berkman LF, Gruenewald TL, Seeman TE. Social integration is associated with fibrinogen concentration in elderly men. Psychosom Med. 2005;67(3):353–8.

217. Loucks EB, Sullivan LM, D'Agostino RBS, Larson MG, Berkman LF, Benjamin EJ. Social networks and inflammatory markers in the Framingham Heart Study. J Biosoc Sci. 2006;38(06):835–42.

218. Kiecolt-Glaser JK, Gouin J-P, Hantsoo L. Close relationships, inflammation, and health. Neurosci Biobehav Rev. 2010;35(1):33–8.

219. Berkman LF. The changing and heterogeneous nature of aging and longevity: a social and biomedical perspective. Annu Rev Gerontol Geriatr. 1988;8:37–68.

220. Geronimus AT, Hicken MT, Pearson JA, Seashols SJ, Brown KL, Cruz TD. Do US black women experience stress-related accelerated biological aging? Hum Nat. 2010;21(1):19–38.

221. Geronimus AT. Black/white differences in the relationship of maternal age to birthweight: a population-based test of the weathering hypothesis. Soc Sci Med. 1996;42(4):589–97.

222. Geronimus AT, Hicken M, Keene D, Bound J. "Weathering" and age patterns of allostatic load scores among blacks and whites in the United States. Am J Public Health. 2006;96(5):826–33.

223. Keene DE, Geronimus AT. "Weathering" HOPE VI: The importance of evaluating the population health impact of public housing demolition and displacement. J Urban Health. 2011;88(3):417–35.

224. Juster RP, McEwen BS, Lupien SJ. Allostatic load biomarkers of chronic stress and impact on health and cognition. Neurosci Biobehav Rev. 2009;35(1):2–16.

225. Lewis TT, Everson-Rose SA, Powell LH, Matthews KA, Brown C, Karavolos K, et al. Chronic exposure to everyday discrimination and coronary artery calcification in African-American women: the SWAN Heart Study. Psychosom Med. 2006;68(3):362–8.

226. Seeman TE, McEwen BS, Rowe JW. Allostatic load as a marker of cumulative biological risk: MacArthur studies of successful aging. PNAS. 2001;98(8):4770–5.

227. Suomi SJ. How gene-environment interactions shape the development of impulsive aggression in rhesus monkeys. In: Stoff DM, Susman EJ, editors. Developmental psychobiology of aggression. New York: Oxford University Press; 2005. pp. 252–70.

228. Suomi SJ. Early determinants of behaviour: evidence from primate studies. British Medical Bulletin. 1997;53(1):170–84.

229. Suomi SJ. Uptight and laid-back monkeys: individual differences in the response to social challenges. In: Brauth SE, Hall WS, Dooling RJ, editors. Plasticity of development. Cambridge, MA: MIT Press; 1991.

230. Provencal N, Suderman MJ, Guillemin C, Massart R, Ruggiero A, Wang D, et al. The signature of maternal rearing in the methylome in rhesus macaque prefrontal cortex and T cells. J Neurosci. 2012;32(44):15626–42.

231. Szyf M, Meaney MJ, Turecki G, Hallet M, Hertzman C, Power C, et al. Epigenetic mechanisms mediating the long-term impact on behavior of the social environment in early life. Amino Acids. 2009;37:16–7.

232. Moss MB, Albert MS. Future directions in the study of aging. In: Albert MS, Moss MB, editors. Geriatric neuropsychology. New York: Guilford Press; 1988. pp. 293–304.

233. Cotman CW. Synaptic plasticity. New York: Guilford Press; 1985.

234. Yan T, Escarce JJ, Liang L-J, Longstreth WT, Merkin SS, Ovbiagele B, et al. Exploring psychosocial pathways between neighbourhood characteristics and stroke in older adults: the cardiovascular health study. Age Ageing. 2013;42(3):391–7.

235. Buchner DM, Beresford SA, Larson EB, LaCroix AZ, Wagner EH. Effects of physical activity on health status in older adults: II. Intervention studies. Annu Rev Public Health. 1992;13:469–88.

236. Emery CF, Gatz M. Psychological and cognitive effects of an exercise program for community-residing older adults. Gerontologist. 1990;30(2):184–8.

237. Wolinsky FD, Stump TE, Clark DO. Antecedents and consequences of physical activity and exercise among older adults. Gerontologist. 1995;35(4):451–62.

238. Davidson RJ, McEwen BS. Social influences on neuroplasticity: stress and interventions to promote well-being. Nat Neurosci. 2012;15(5):689–95.

239. Lavretsky H, Epel ES, Siddarth P, Nazarian N, Cyr NS, Khalsa DS, et al. A pilot study of yogic meditation for family dementia caregivers with depressive symptoms: effects on mental health, cognition, and telomerase activity. Int J Geriatr Psychiatry. 2012;28(1):57–65.

240. Vetencourt JFM, Sale A, Viegi A, Baroncelli L, De Pasquale R, O'Leary OF, et al. The antidepressant fluoxetine restores plasticity in the adult visual cortex. Science. 2008;320(5874):385–8.

241. Orth-Gomer K, Unden AL. The measurement of social support in population surveys. Soc Sci Med. 1987;24(1):83–94.

242. Donaldson CA, Ware JE. The qualification of social contacts and resources. Santa Monica, CA: Rand Corporation; 1982.

243. Berkman LF, Breslow L. Health and ways of living: the Alameda County study. New York: Oxford University Press; 1983.

244. Boissevain J. Friends of friends: networks, manipulators and coalitions. New York: St Martins Press; 1974.

245. Glass TA, Mendes de Leon CF, Seeman TE, Berkman LF. Beyond single indicators of social networks: a LISREL analysis of social ties among the elderly. Soc Sci Med. 1997;44(10):1503–17.

246. Mitchell JC. The concept and use of social networks. In: Mitchell JC, editor. Social networks in urban situations: analyses of personal relationships in Central African towns. Manchester, UK: Manchester University Press; 1969. pp. 1–50.

247. Burt RS. Toward a structural theory of action. New York: Academic Press; 1982.

248. Cohen S, Hoberman HM. Positive events and social supports as buffers of life change stress. J Appl Social Pyschol. 1983;13(2):99–125.

249. Procidano ME, Heller K. Measures of perceived social support from friends and from family: three validation studies. Am J Community Psychol. 1983;11(1):1–24.

250. Barrera M Jr., Sandler IN, Ramsay TB. Preliminary development of a scale of social support: studies on college students. Am J Community Psychol. 1981;9(4):435–47.

251. Sherbourne CD, Stewart AL. The MOS social support survey. Soc Sci Med. 1991;32(6):705–14.

252. Blumenthal JA, Burg MM, Barefoot J, Williams RB, Haney T, Zimet G. Social support, type A behavior, and coronary artery disease. Psychosom Med. 1987;49(4):331–40.

253. Blumenthal JA, Babyak MA, Moore KA, Craighead WE, Herman S, Khatri P, et al. Effects of exercise training on older patients with major depression. Arch Intern Med. 1999;159(19):2349–56.

254. Rozanski A, Blumenthal JA, Kaplan J. Impact of psychological factors on the pathogenesis of cardiovascular disease and implications for therapy. Circulation. 1999;99(16):2192–217.

255. Mitchell PH, Powell L, Blumenthal J, Norten J, Ironson G, Pitula CR, et al. A short social support measure for patients recovering from myocardial infarction: the ENRICHD Social Support Inventory. J Cardiopulm Rehabil. 2003;23(6):398–403.

256. Henderson S, Duncan-Jones P, Byrne DG, Scott R. Measuring social relationships: the interview schedule for social interaction. Psychol Med. 1980;10(4):723–34.

257. Unden AL, Orth-Gomer K. Development of a survey method to measure social support in population studies. Stockholm: Karolinska Institute; 1984. Report No.: Stress Research Report No. 178.

258. Lubben J, Blozik E, Gillmann G, Iliffe S, Renteln Kruse von W, Beck JC, et al. Performance of an abbreviated version of the Lubben Social Network Scale among three European community-dwelling older adult populations. Gerontologist. 2006;46(4):503–13.

259. Luo Y, Hawkley LC, Waite LJ, Cacioppo JT. Loneliness, health, and mortality in old age: a national longitudinal study. Soc Sci Med. 2012;74(6):907–14.

260. Hawkley LC, Cole SW, Capitanio JP, Norman GJ, Cacioppo JT. Effects of social isolation on glucocorticoid regulation in social mammals. Horm Behav. 2012;62(3):314–23.

261. Hawkley LC, Hughes ME, Waite LJ, Masi CM, Thisted RA, Cacioppo JT. From social structural factors to perceptions of relationship quality and loneliness: the Chicago health, aging, and social relations study. J Gerontol B Psychol Sci Soc Sci. 2008;63(6):S375–84.

262. Hawkley LC, Masi CM, Berry JD, Cacioppo JT. Loneliness is a unique predictor of age-related differences in systolic blood pressure. Psychol Aging. 2006;21(1):152–64.

263. Hughes ME, Waite LJ, Hawkley LC, Cacioppo JT. A short scale for measuring loneliness in large surveys: results from two population-based studies. Res Aging. 2004;26(6):655–72.

264. Cacioppo JT, Norris CJ, Decety J, Monteleone G, Nusbaum H. In the eye of the beholder: individual differences in perceived social isolation predict regional brain activation to social stimuli. J Cogn Neurosci. 2009;21(1):83–92.

265. Cacioppo JT, Hawkley LC, Thisted RA. Perceived social isolation makes me sad: 5-year cross-lagged analyses of loneliness and depressive symptomatology in the Chicago Health, Aging, and Social Relations Study. Psychol Aging. 2010;25(2):453–63.

266. Kiecolt-Glaser JK, Malarkey WB, Chee M, Newton T, Cacioppo JT, Mao HY, et al. Negative behavior during marital conflict is associated with immunological down-regulation. Psychosom Med. 1993;55(5):395–409.

267. Seeman TE, Singer BH, Ryff CD, Love GD, Levy-Storms L. Social relationships, gender, and allostatic load across two age cohorts. Psychosom Med. 2002;64(3):395–406.

268. Brooks KP, Dunkel Schetter C. Social negativity and health: conceptual and measurement issues. Soc Personal Psychol Compass. 2011;5(11):904–18.

269. Smith TW, Ruiz JM, Uchino BN. Mental activation of supportive ties, hostility, and cardiovascular reactivity to laboratory stress in young men and women. Health Psychol. 2004;23(5):476–85.

270. Schuster TL, Kessler RC, Aseltine RH Jr. Supportive interactions, negative interactions, and depressed mood. Am J Community Psychol. 1990;18(3):423–38.

271. Anderson D, Deshaies G, Jobin J. Social support, social networks and coronary artery disease rehabilitation: a review. Can J Cardiol. 1996;12(8):739–44.

272. Greenwood DC, Muir KR, Packham CJ, Madeley RJ. Coronary heart disease: a review of the role of psychosocial stress and social support. J Public Health Med. 1996;18(2):221–31.

273. Helgeson VS, Cohen S. Social support and adjustment to cancer: reconciling descriptive, correlational, and intervention research. Health Psychol. 1996;15(2):135–48.

274. Seeman TE. Social ties and health: the benefits of social integration. Ann Epidemiol. 1996;6(5):442–51.

275. Eriksen W. The role of social support in the pathogenesis of coronary heart disease: a literature review. Fam Pract. 1994;11(2):201–9.

276. Riley AW, Coiro MJ, Broitman M, Colantuoni E, Hurley KM, Bandeen-Roche K, et al. Mental health of children of low-income depressed mothers: influences of parenting, family environment, and raters. Psychiatr Serv. 2009;60(3):329–36.

277. Masi CM, Chen H-Y, Hawkley LC, Cacioppo JT. A meta-analysis of interventions to reduce loneliness. Pers Soc Psychol Rev. 2011;15(3):219–66.

278. Zunzunegui M-V, Alvarado BE, Del Ser T, Otero A. Social networks, social integration, and social engagement determine cognitive decline in community-dwelling Spanish older adults. J Gerontol B Psychol Sci Soc Sci. 2003;58(2):S93–S100.

279. Kumar S, Calvo R, Avendano M, Sivaramakrishnan K, Berkman LF. Social support, volunteering and health around the world: cross-national evidence from 139 countries. Soc Sci Med. 2012;74(5):696–706.

280. Schoenbach VJ, Kaplan BH, Fredman L, Kleinbaum DG. Social ties and mortality in Evans County, Georgia. Am J Epidemiol. 1986;123(4):577–91.

281. Wellin B. May God's blessings now and ever after rest upon the work in this association: interview by Viveka Holmertz and Inger Lernevall. Vårdfacket. 1985;9(10):20–1.

282. Kaplan GA, Salonen JT, Cohen RD, Brand RJ, Syme SL, Puska P. Social connections and mortality from all causes and from cardiovascular disease: prospective evidence from eastern Finland. Am J Epidemiol. 1988;128(2):370–80.

283. Vogt T. Social networks as predictors of ischemic heart disease, cancer, stroke and hypertension: incidence, survival and mortality. J Clin Epidemiol. 1992;45(6):659–66.

284. Kawachi I, Colditz GA, Ascherio A, Rimm EB, Giovannucci E, Stampfer MJ, et al. A prospective study of social networks in relation to total mortality and cardiovascular disease in men in the USA. J Epidemiol Community Health. 1996;50(3):245–51.

285. Penninx BWJH, van Tilburg T, Kriegsman DMW, Deeg DJH, Boeke AJP, van Eijk JTM. Effects of social support and personal coping resources on mortality in older age: the Longitudinal Aging Study Amsterdam. Am J Epidemiol. 1997;146(6):510–9.

286. Sugisawa H, Liang J, Liu X. Social networks, social support, and mortality among older people in Japan. J Gerontol. 1994;49(1):S3–S13.

287. Berkman LF, Melchior M, Chastang JF, Niedhammer I, Leclerc A, Goldberg M. Social integration and mortality: a prospective study of French employees of Electricity of France–Gas of France: the GAZEL Cohort. Am J Epidemiol. 2004;159(2):167–74.

288. Orth-Gomer K, Rosengren A, Wilhelmsen L. Lack of social support and incidence of coronary heart disease in middle-aged Swedish men. Psychosom Med. 1993;55(1):37–43.

289. Ruberman W, Weinblatt E, Goldberg JD, Chaudhary BS. Psychosocial influences on mortality after myocardial infarction. N Engl J Med. 1984;311(9):552–9.

290. Orth-Gomer K, Johnson JV. Social network interaction and mortality. J Chronic Dis. 1987;40(10):949–57.

291. Williams RB. Prognostic importance of social and economic resources among medically treated patients with angiographically documented coronary artery disease. JAMA. 1992;267(4):520.

292. Case RB, Moss AJ, Case N, McDermott M, Eberly S. Living alone after myocardial infarction: impact on prognosis. JAMA. 1992;267(4):515–9.

293. Berkman LF, Leo-Summers L, Horwitz RI. Emotional support and survival after myocardial infarction: A prospective, population-based study of the elderly. Ann Intern Med. American College of Physicians; 1992;117(12):1003–9.

294. Oxman TE, Freeman DH, Manheimer ED. Lack of social participation or religious strength and comfort as risk factors for death after cardiac surgery in the elderly. Psychosom Med. 1995;57(1):5–15.

295. Farmer IP, Meyer PS, Ramsey DJ, Goff DC, Wear ML, Labarthe DR, et al. Higher levels of social support predict greater survival following acute myocardial infarction: the Corpus Christi Heart Project. Behav Med. 1996;22(2):59–66.

296. Krumholz HM, Butler J, Miller J, Vaccarino V, Williams CS, Mendes de Leon CF, et al. Prognostic importance of emotional support for elderly patients hospitalized with heart failure. Circulation. 1998;97(10):958–64.

297. Chen YT, Vaccarino V, Williams CS, Butler J, Berkman LF, Krumholz HM. Risk factors for heart failure in the elderly: a prospective community-based study. Am J Med. 1999;106(6):605–12.

298. Morris P, Robinson RG. Association of depression with 10-year poststroke mortality. Am J Psychiat. 1993;150(1):124–9.

299. Strogatz DS, Croft JB, James SA, Keenan NL, Browning SR, Garrett JM, et al. Social support, stress, and blood pressure in black adults. Epidemiology. 1997;8(5):482.

300. Evans RL, Bishop DS, Matlock AL, Stranahan S, Halar EM, Noonan WC. Prestroke family interaction as a predictor of stroke outcome. Arch Phys Med Rehabil. 1987;68(8):508–12.

301. Friedland J, McColl M. Social support and psychosocial dysfunction after stroke: buffering effects in a community sample. Arch Phys Med Rehabil. 1987;68(8):475–80.

302. Glass TA, Matchar DB, Belyea M, Feussner JR. Impact of social support on outcome in first stroke. Stroke. 1993;24(1):64–70.

303. McLeroy KR, DeVellis R, DeVellis B, Kaplan B, Toole J. Social support and physical recovery in a stroke population. J Soc Pers Relat. 1984;1(4):395–413.

304. Robertson EK, Suinn RM. The determination of rate of progress of stroke patients through empathy measures of patient and family. J Psychosom Res. 1968;12(3):189–91.

305. Angeleri F, Angeleri VA, Foschi N, Giaquinto S, Nolfe G. The influence of depression, social activity, and family stress on functional outcome after stroke. Stroke. 1993;24(10):1478–83.

306. Evans RL, Connis RT, Bishop DS, Hendricks RD, Haselkorn JK. Stroke: a family dilemma. Disabil Rehabil. 1994;16(3):110–8.

307. King RB. Quality of life after stroke. Stroke. 1996;27(9):1467–72.

308. Hyman MD. Social isolation and performance in rehabilitation. J Chronic Dis. 1972;25(2):85–97.

309. Kishi Y, Kosier JT, Robinson RG. Suicidal plans in patients with acute stroke. J Nerv Ment Dis. 1996;184(5):274–80.

310. Colantonio A, Kasl SV, Ostfeld AM, Berkman LF. Psychosocial predictors of stroke outcomes in an elderly population. J Gerontol. 1993;48(5):S261–8.

311. Lehmann JF, DeLateur BJ, Fowler RS Jr, Warren CG, Arnhold A, Schertzer G, et al. Stroke rehabilitation: outcome and prediction. Arch Phys Med Rehabil. 1975;56(9):383–9.

312. Brosseau L, Potvin L, Philippe P, Boulanger YL. Post-stroke inpatient rehabilitation: II. Predicting discharge disposition. Am J Phys Med Rehabil. 1996;75(6):431–6.

313. Kwakkel G, Wagenaar RC, Kollen BJ, Lankhorst GJ. Predicting disability in stroke—a critical review of the literature. Age Ageing. 1996;25(6):479–89.

314. Evans RL, Matlock AL, Bishop DS, Stranahan S, Pederson C. Family intervention after stroke: does counseling or education help? Stroke. 1988;19(10):1243–9.

315. Oldenburg B, Perkins RJ, Andrews G. Controlled trial of psychological intervention in myocardial infarction. J Consult Clin Psychol. 1985;53(6):852–9.

316. Spiegel D, Kraemer HC, Bloom J, Gottheil E. Effect of psychosocial treatment on survival of patients with metastatic breast cancer. Lancet. 1989;334(8668):888–91.

317. Bassuk SS, Glass TA, Berkman LF. Social disengagement and incident cognitive decline in community-dwelling elderly persons. Ann Intern Med. 1999;131(3):165–73.

318. Fratiglioni L, Paillard-Borg S, Winblad B. An active and socially integrated lifestyle in late life might protect against dementia. Lancet Neurology. 2004;3(6):343–53.

319. Lobo A, Launer LJ, Fratiglioni L, Anderson K, DiCarlo A, Breteler MMB, et al. Prevalence of dementia and major subtypes in Europe: a collaborative study of population-based cohorts. Neurology. 2000;54(11):S4–S9.

320. Karp A, Paillard-Borg S, Wang H-X, Silverstein M, Winblad B, Fratiglioni L. Mental, physical and social components in leisure activities equally contribute to decrease dementia risk. Dement Geriatr Cogn Disord. 2006;21(2):65–73.

321. Barnes LL, Mendes de Leon CF, Wilson RS, Bienias JL, Evans DA. Social resources and cognitive decline in a population of older African Americans and whites. Neurology. 2004;63(12):2322–6.

322. Glei DA. Participating in social activities helps preserve cognitive function: an analysis of a longitudinal, population-based study of the elderly. Int J Epidemiol. 2005;34(4):864–71.

323. Saczynski JS, Pfeifer LA, Masaki K, Korf ESC, Laurin D, White L, et al. The effect of social engagement on incident dementia. Am J Epidemiol. 2006;163(5).

324. Ertel KA, Glymour MM, Berkman LF. Effects of social integration on preserving memory function in a nationally representative US elderly population. Am J Public Health. 2008;98(7):1215–20.

325. Bosma H, van Boxtel MPJ, Ponds RWHM, Jelicic M, Houx P, Metsemakers J, et al. Engaged lifestyle and cognitive function in middle and old-aged, non-demented persons: a reciprocal association? Z Gerontol Geriatr. 2002;35(6):575–81.

326. Hultsch DF, Hertzog C, Small BJ, Dixon RA. Use it or lose it: engaged lifestyle as a buffer of cognitive decline in aging? Psychol Aging. 1999;14(2):245.

327. Carlson MC, Erickson KI, Kramer AF, Voss MW, Bolea N, Mielke M, et al. Evidence for neurocognitive plasticity in at-risk older adults: the Experience Corps program. J Gerontol. 2009;64A(12):1275–82.

328. Entwisle B, Faust K, Rindfuss RR, Kaneda T. Networks and contexts: variation in the structure of social ties. Am J Sociol. 2007;112(5):1495–533.

329. Bärnighausen T, Tanser F, Gqwede Z, Mbizana C, Herbst K, Newell M-L. High HIV incidence in a community with high HIV prevalence in rural South Africa: findings from a prospective population-based study. AIDS. 2008;22(1):139–44.

330. Bärnighausen T, Hosegood V, Timaeus IM, Newell M-L. The socioeconomic determinants of HIV incidence: evidence from a longitudinal, population-based study in rural South Africa. AIDS. 2007;21(Suppl 7):S29–S38.

331. Eaton JW, Johnson LF, Salomon JA, Bärnighausen T, Bendavid E, Bershteyn A, et al. HIV treatment as prevention: systematic comparison of mathematical models of the potential impact of antiretroviral therapy on HIV incidence in South Africa. PLoS Med. 2012;9(7):e1001245.

332. Neaigus A, Friedman SR, Curtis R, Jarlais Des DC, Terry Furst R, Jose B, et al. The relevance of drug injectors' social and risk networks for understanding and preventing HIV infection. Soc Sci Med. 1994;38(1):67–78.

333. Latkin C, Mandell W, Oziemkowska M, Celentano D, Vlahov D, Ensminger M, et al. Using social network analysis to study patterns of drug use among urban drug users at high risk for HIV/AIDS. Drug Alcohol Depend. 1995;38(1):1–9.

334. Newcomb MD, Bentler PM. Impact of adolescent drug use and social support on problems of young adults: a longitudinal study. J Abnorm Psychol. 1988;97(1):64–75.

335. Zapka JG, Stoddard AM, McCusker J. Social network, support and influence: relationships with drug use and protective AIDS behavior. AIDS Educ Prev. 1993;5(4):352–66.

336. Friedman SR, Neaigus A, Jarlais Des DC, Sotheran JL, Woods J, Sufian M, et al. Social intervention against AIDS among injecting drug users. Addiction. 1992;87(3):393–404.

337. Kelly JA, Murphy DA, Sikkema KJ, Kalichman SC. Psychological interventions to prevent HIV infection are urgently needed: new priorities for behavioral research in the second decade of AIDS. Am Psychol. 1993;48(10):1023–34.

338. Kelly JA, St Lawrence JS, Diaz YE, Stevenson LY, Hauth AC, Brasfield TL, et al. HIV risk behavior reduction following intervention with key opinion leaders of population: an experimental analysis. Am J Public Health. 1991;81(2):168–71.

339. Theorell T, Blomkvist V, Jonsson H, Schulman S, Berntorp E, Stigendal L. Social support and the development of immune function in human immunodeficiency virus infection. Psychosom Med. 1995;57(1):32–6.

340. Wallace R. Traveling waves of HIV infection on a low dimensional "socio-geographic" network. Soc Sci Med. 1991;32(7):847–52.

341. Hunt CW. Migrant labor and sexually transmitted disease: AIDS in Africa. J Health Soc Behav. 1989;30(4):353–73.

342. Selye H. The stress of life. New York: McGraw-Hill; 1956.

343. Esterling BA, Kiecolt-Glaser JK, Glaser R. Psychosocial modulation of cytokine-induced natural killer cell activity in older adults. Psychosom Med. 1996;58(3):264–72.

344. Glaser R, Kiecolt-Glaser JK, Bonneau RH, Malarkey W, Kennedy S, Hughes J. Stress-induced modulation of the immune response to recombinant hepatitis B vaccine. Psychosom Med. 1992;54(1):22–9.

345. Marsden PV. Models and methods in social network analysis. In: Carrington PJ, Scott JJ, Wasserman S, editors. Models and methods in social network analysis New York: Cambridge University Press; 2005. pp. 8–30.

346. Marsden PV. Network methods in social epidemiology. In: Oakes JM, Kaufman JS, editors. Methods in social epidemiology. Hoboken, NJ: John Wiley & Sons; 2006. pp. 267–86.

347. Marsden PV. Survey methods for network data. In: Scott JJ, Carrington PJ, editors, The SAGE handbook of social network analysis. Thousand Oaks, CA: Sage Publications; 2011. pp. 370–88.

348. Hollstein B. Qualitative approaches. In Scott J, Carrington PJ, editors. The SAGE handbook of social network analysis. Thousand Oaks, CA: Sage Publications; 2011. pp. 404–16.

349. Keating NL, Ayanian JZ, Cleary PD, Marsden PV. Factors affecting influential discussions among physicians: a social network analysis of a primary care practice. J Gen Intern Med. 2007;22(6):794–8.

350. Davis JA, Smith TW, Marsden PV. General social surveys, 1972–2006 [cumulative file]. Inter-University Consortium for Political and Social Research, 2007.

351. Lin N, Fu YC, Hsung RM. The position generator: measurement techniques for investigations of social capital. In: Lin N, Cook KS, Burt RS, editors. Social capital: theory and research. New York: Aldine de Gruyter; 2001. pp. 57–81.

352. Lin N, Simeone RS, Ensel WM, Kuo W. Social support, stressful life events, and illness: a model and an empirical test. J Health Soc Behav. 1979;20(2):108–19.

353. Dubow EF, Ullman DG. (1989). Assessing social support in elementary school children: the survey of children's social support. J Clin Child Psychol. 18(1):52–64.

354. Dehle C, Larsen D, Landers JE. Social support in marriage. Am J Fam Ther. 2001;29(4):307–24.

355. Lubben JE. Assessing social networks among elderly populations. *Fam Community Health.* 1988;11(3): 42–52.

356. Oh K, Oh K-O, Lee S-J, Kim J-A, Jeong C-J, Kim H-R, et al. Psychometric evaluation of the Korean Social Support Questionnaire. J Korean Acad Nurs. 2008;38(6):881.

357. Norbeck JS, Lindsey AM, Carrieri VL. The development of an instrument to measure social support. Nurs Res. 1981;30(5):264–69.

358. Ong ASJ, Ward C. The construction and validation of a social support measure for sojourners: the Index of Sojourner Social Support (ISSS) Scale. J Cross Cult Psychol. 2005;36(6):637–61.

359. Zimet G, Dahlem N, Zimet S, Farley G. The multidimensional scale of perceived social support. J Pers Assess. 1988;52:30–41.

360. Stansfeld S, Marmot M. Deriving a survey measure of social support: the reliability and validity of the close persons questionnaire. Soc Sci Med. 1992;35(8):1027–35.

361. Ryff CD. Happiness is everything, or is it? Explorations on the meaning of psychological well-being. J Pers Soc Psychol. 1989;57:1069–81.

362. Vinokur AD, Price RH, Caplan RD. Hard times and hurtful partners: how financial strain affects depression and relationship satisfaction of unemployed persons and their spouses. J Pers Soc Psychol. 1996;71(1):166–179.

363. Lakey B, Tardiff TA, Drew JB. Negative social interactions: assessment and relations to social support, cognition, and psychological distress. J Soc Clin Psychol. 1994;13(1):42–62.

364. Repetti RL, Wood J. Effects of daily stress at work on mothers' interactions with preschoolers. J Fam Psychol. 1997;11(1):90–108.

365. Repetti RL. Short-term effects of occupational stressors on daily mood and health complaints. Health Psychol. 1993;12:125–31.

366. Repetti RL. Effects of daily workload on subsequent behavior during marital interaction: the roles of social withdrawal and spouse support. J Pers Soc Psychol.1989;57(4):651–9.

367. Story LB, Repetti R. Daily occupational stressors and marital behavior. J Fam Psychol. 2006;20(4):690–700.

368. Moos R, Moos B. Family environment scale. Palo Alto, CA: Consulting Psychologists Press; 1981.

369. Spanier GB. Measuring dyadic adjustment: new scales for assessing the quality of marriage and similar dyads. J Marriage Fam. 1976;38(1):15–28.

370. Russell D, Peplau LA, Cutrona CE. The revised UCLA Loneliness Scale: concurrent and discriminant validity evidence. J Pers Soc Psychol. 1980;39(3):472–80.

371. de Jong Gierveld J, van Tilburg T. Living arrangements of older adults in the Netherlands and Italy: coresidence values and behaviour and their consequences for loneliness. J Cross Cult Gerontol. 1999;14(1):1–24.

372. Vinconzi H, Grabosky, F. Measuring the emotional/social aspects of loneliness and isolation. J Pers Soc Psychol. 1987;2(2):257–270.

373. DiTommaso E, Brannen C, Best LA. Measurement and validity characteristics of the short version of the social and emotional loneliness scale for adults. Educ Psychol Meas. 2004;64(1):99–119.

374. Asher SR, Hymel S, Renshaw PD. Loneliness in children. Child Dev. 1984;55:1456–64.

375. Lawton MP. The Philadelphia Geriatric Center Morale Scale: a revision. J Gerontol. 1975;30(1):85–9.

376. Paloutzian RF, Ellison CW. Loneliness, spiritual well-being and quality of life. In: Peplau LA, Perlman D, editors. Loneliness: a sourcebook of current theory, research and therapy. New York: John Wiley & Sons; 1982. pp. 224–7.

377 . Duke University. Multidimensional functional assessment: the OARS. Durham, NC: Center for the Study of Aging and Human Development, Duke University; 1978.

378. Morrow-Howell N, Becker-Kemppainen S, Judy L. Evaluating an intervention for the elderly at increased risk of suicide. Res Social Work Prac. 1998;8(1):28–46.

379. Chadsey-Rusch J, DeStefano L, O'Reilly M, Gonzalez P, Collet-Klingenberg L. Assessing the loneliness of workers with mental retardation. Ment Retard. 1992;30(2):85–92.

380. Batagelj V, Mrvar A, de Nooy W. Exploratory social network analysis with Pajek. New York: Cambridge University Press; 2008.

381. Batagelj V, Mrvar A. PAJEK, ver. 91. 2001. http://vlado.fmf.uni-lj.si/pub/networks/pajek/

382. Cyram. NetMiner II, ver. 2.5.0a. Seoul: Cyram; 2004.

383. Borgatti SP. NetDraw Software for network visualization. Lexington, KY: Analytic Technologies; 2002.

384. Borgatti SP, Everett MG, Freeman LC. UCINET V for Windows: software for social network analysis, ver. 5.2.0.1. Natick, MA: Analytic Technologies; 1999.

385. Goodreau SM, Handcock MS, Hunter DR, Butts CT, Morris M. A statnet tutorial. J Stat Softw. 2008;24(9):1–27.

386. Handcock MS, Hunter DR, Butts CT, Goodreau SM, Morris M. statnet: software tools for the representation, visualization, analysis and simulation of network data. J Stat Softw. 2008;24(1):1548–7660.

387. Hunter DR, Handcock MS, Butts CT, Goodreau SM, Morris M. ergm: a package to fit, simulate and diagnose exponential-family models for networks. J Stat Softw. 2008;24(3):nihpa54860.

CHAPTER 8

SOCIAL CAPITAL, SOCIAL COHESION, AND HEALTH

Ichiro Kawachi and Lisa F. Berkman

TWO VIGNETTES

In the early morning of Tuesday, January 17, 1995, a devastating earthquake struck the city of Kobe, the fifth-largest city in Japan. Over 150,000 buildings collapsed, and more than 6,000 people lost their lives. As told by Daniel Aldrich (1), the disaster uncovered striking differences in the preparedness and speed of response across the city's neighborhoods. In the neighborhood of Mano—long known and studied for its vibrant community organizations[1]—the residents spontaneously formed bucket brigades to put out fires, while in nearby neighborhoods, residents watched helplessly as their neighbors' homes burned to the ground (1). In the immediate aftermath of the earthquake, local neighborhood associations assisted in rescue operations, helped to evacuate homeless residents to nearby schools, established community kitchens, and organized night watchmen to guard abandoned property (2). During the reconstruction phase, the same organizations facilitated the inspection of damaged buildings; they published weekly newsletters to keep community residents informed; and they helped to oversee the process of retrofitting damaged houses. Residents of neighborhoods like Mano were also quicker than the rest of the city to establish new organizations that lobbied effectively to bring resources to local businesses for reconstruction. In the Kobe earthquake—as well as in countless other disasters throughout the world (1, 3, 4)—what makes the decisive difference in the human toll turns out to be not just the physical dimensions of the disaster (e.g., the magnitude of the earthquake on the Richter scale), but also the human and *social* dimension.

1 Known as *machizukuri*—or literally "neighborhood-building"—associations, which were originally organized by the residents to deal with problems such as pollution from nearby factories. Later, their mandate expanded to deal with additional concerns of residents, such as improving parks and recreational space, preventing crime, and so forth.

Indian society is periodically wracked by ethnic conflict between Hindu and Muslim populations. Yet, as Varshney (5) has shown, the proportion of Muslims versus Hindus in an urban area cannot adequately account for why some cities suffer from a long history of ethnic violence while others are successful in maintaining harmony. What is the secret ingredient of peace? Ashutosh Varshney's (5) answer is that cities that are able to maintain the peace are often characterized by the presence of ethnically integrated civic organizations—for example, business groups, trade unions, and reading circles in local libraries—that include both Muslims and Hindus among their members. These organizations, Varshney conjectures, are effective means of keeping the channels of communication open across ethnic groups, and serve to stop rumors that agitators often spread within their communities in order to incite the population to riot.

What is the mysterious community characteristic shared by these two apparently unrelated vignettes? In both scenarios, elements of the community structure—taking the form of preexisting community-based organizations—helped the residents to recover quickly from natural disasters (in Japan) and to avoid ethnic conflict (in India). The network of connections between residents constitutes a part of the *social capital* of the neighborhoods where they live.

DEFINITION OF SOCIAL CAPITAL

A simple definition of social capital is the "resources that are accessed by individuals as a result of their membership of a network or a group." In fact, the definition of social capital is anything but "simple." A dizzying number of variations have been offered in the social sciences. Perhaps this is inevitable, given that so many disciplines have weighed in on the concept, ranging from sociology (6, 7), to economics (8), to political science (9), and population health (10). It is part of the nature of interdisciplinary discourse that the definitions of constructs become fuzzy and blurred. Nevertheless, most definitions of social capital emphasize two features: *(1)* it is a resource; and *(2)* it is generated through social connections.

In his essay about "the forms of capital," Bourdieu (6) famously argued that the use of the term "capital" ought not to be restricted to financial capital alone.[2] In other words, in quotidian conversation, capital is equated with money, but it is a mistake to stop there. Capital can refer to any stock of goods or resources. For example, economists refer to the stock of knowledge and skills acquired over a lifetime of learning as "human capital." Bourdieu (6) referred to the acquisition of certain habits (e.g., going to museums and concerts), preferences, or styles of speech and dress as "cultural capital," which individuals use to express their symbolic status in society. "Social capital" expresses the idea that there are tangible resources embedded in social relationships, available for members to access. That is, when we hang out with our friends, we may be simply having a good time, but at the same time, our network connections are also available for us to draw on material and psychological resources. For this reason, social capital is sometimes referred to as "network capital."

2 "It is in fact impossible to account for the structure and functioning of the social world unless one reintroduces capital in all its forms and not solely in one form recognized by economic theory. Economic theory has allowed to be foisted upon it a definition of the economy of practices which is the historical invention of capitalism; and by reducing the universe of exchanges to mercantile exchange" (Bourdieu, 1986, p. 241).

In formal economic theory, capital has two characteristics: *(1)* its creation entails a sacrifice of present consumption for future benefit; and *(2)* it enhances the productivity of other factors of production (11). Following this reasoning, education is a form of capital, since: *(1)* people sacrifice fun (as well as earnings) to stay in school to get educated, and *(2)* schooling enhances the productivity of other factors of production (such as operating sophisticated widgets) (see Chapter 2). By contrast, eminent economists such as Kenneth Arrow (12) have argued that "social capital" fails the formal test of the definition since there is no sacrifice or purposive "investment" in network connections for future gain. While we agree that most people do not socialize for purely instrumental ends (we hang out with friends because they are fun), we note that economists are also fond of pointing out that there are opportunity costs of time. When people volunteer in civic organizations—for example, local residents' associations—they in fact sacrifice present consumption (after all it would be so much more fun to drink beer and watch football on TV) in favor of building up stocks of community social capital.

THEORETICAL PATHWAYS LINKING SOCIAL CAPITAL TO HEALTH

The pathways linking social capital to health outcomes vary according to the level of analysis. Analyzed at the *individual* level, social capital refers to resources accessed by the individual through their ego-centered networks. Examples of health-relevant resources include the acquisition of useful information (e.g., news about where to get a free flu shot), the receipt of instrumental support (e.g., cash loans), and social reinforcement (e.g., the exchange of affective support). Analyzed at the individual level, "social capital" can be sometimes difficult to distinguish from the concept of "social support" (discussed in Chapter 7). However, one important distinction is that social support—at least as it is commonly assessed in social epidemiology via ego-centered network measures (see Chapter 7)—derives from close, strong ties that individuals maintain. Individual social capital can also be accessed from close strong ties but it also comes much more from *weak*, acquaintance ties, as measured by instruments such as the Resource Generator (see section below on measurement of social capital). In this sense, individual social capital may be seen to represent the diversity (i.e., weak ties, bridging relationships across groups) in someone's network in contrast to support per se. For example, individual network social capital, particularly the diversity component, appears to be protective against smoking relapse and hypertension even after adjusting for the individual's strong core ties.[3] In other words, there appears to be something beneficial at the individual level about having diverse networks beyond the ability to recruit social support through intimate ties (13, 14).

When we turn to social capital as a *group*-level construct, we are treating it as a property of the whole social network—for example, a network of connected residents in a community—which can bring benefits to individuals embedded in it. When the analysis is pushed out to the

3 Spencer Moore, Queens University: personal communication. We are indebted to Dr. Moore for the insights covered in this section.

group level, social capital is associated with a set of "emergent" properties (15). Three group-level mechanisms merit particular attention as being potentially relevant to health outcomes: *(1)* social contagion, *(2)* informal social control, and *(3)* collective efficacy.

Social contagion references the notion that behaviors spread more quickly through a tightly knit social network. In network terminology, the greater the transitivity between members of a network (i.e., the more saturated the social connections between individuals in a network), the more paths there are for members to influence the behavior of others in the network. Behaviors can spread through a network through the diffusion of information or through the transmission of behavioral norms. Sometimes the behavior that spreads via the network can be deleterious to health—for example, the spread of obesity through a social network (16)—but at other times, the behavior can be health-promoting, for example, the spread of smoking cessation (17). In the Framingham Offspring Study, Christakis and Fowler (17) found that smoking cessation behavior obeys the "three degrees of influence" rule, that is, everything we do or say ripples through our social network up to three degrees of separation away. Thus, when an index individual stops smoking, it increases the odds that his immediate friends (one degree of separation away) will stop smoking by 60%; but it also raises his friends' friends' odds of stopping smoking by 20% (two degrees away), as well as his friends' friends' friends' odds of quitting by 10% (three degrees) (17). Tellingly, during the three-decade follow-up of the cohort, the few remaining diehard smokers ended up being progressively pushed out toward the periphery of the social network, that is, they found themselves becoming progressively ostracized from their social contacts. What is noteworthy about the "three degrees of influence rule" is that we are dealing with a property of the group. Stated another way, all of us (by definition) know who our friends are, but we may not be familiar with all of our friends' friends (at two degrees), and the chances are quite slim that we know our friends' friends' friends. The results of Christakis and Fowler's (17) study imply that we may be influenced by the behavior of people *whom we do not even know*; yet by virtue of our membership in a network, we may still "benefit" from the rippling contagion triggered by the behavior of others in a distant part of the network. If this account of behavioral contagion is accurate,[4] a corollary is that we expect to observe faster diffusion of behaviors in more cohesive (i.e., higher social capital) networks.

Informal social control refers to the ability of adults in a community to maintain social order, that is, to step in and intervene when they witness deviant behavior by others. The concept originated in criminology to explain community variations in the occurrence of vandalism and delinquency (18). A cohesive community is one in which residents can rely on its adults—not just the parents or formal agents of law enforcement—to step in to intervene when they witness youth loitering on the streets or engaging in unlawful behavior. The likelihood of this type of informal policing rises when there is *network closure*, that is, the adults in a community are socially connected to each other. Although informal social control was originally developed to explain the ability of the community to suppress crime, it is equally applicable and relevant to the prevention of a range of health-related behaviors including underage smoking, drinking, or drug abuse.[5] Whenever a

4 There have been critics of Christakis and Fowler, for example, Cohen-Cole and Fletcher (2008) (73) and Lyons (2011) (74)—see Chapter 7 for more detailed discussion.

5 For example, in the Tokyo neighborhood where the author (IK) grew up in the 1970s, cigarettes were ubiquitously available from the vending machines on every street corner. Many a time he and his classmates were tempted to spend their pocket money to sneak a packet of cigarettes on their way home from school, yet they never actually dared to do it because they knew that their mothers would find out about it before they came home; some parent in the neighborhood would have inevitably called ahead and tattled on them.

parent relies on her neighbors to police the behaviors of their children when they are not looking, she is benefiting from the network to which she belongs. Stated another way, informal control is a collective characteristic of the group.

Collective efficacy is the group-level analog of the concept of self-efficacy, that is, it refers to the ability of the collective to mobilize to undertake collective action (18). In the Kobe earthquake vignette that opened the chapter, neighborhoods with a higher density of civic organizations predating the disaster were better prepared and quicker to get back on their feet (1). When residents of a community are connected to each other through civic and voluntary associations, mobilizing in an emergency happens faster. When a collective problem arises, many (perhaps most) of us would prefer to sit back and let others do the work; this is known as the free-rider problem. Why then do people volunteer their efforts, for example to clean up the debris after an earthquake? One reason is that they may be already connected to each other via existing community-based organizations. Free-riding in this context risks damage to one's reputation as well as social sanctions (i.e., ostracism). The threat of sanction by other members of the group may be sufficient in this case to mobilize collective action. The extent of civic participation in a community is therefore a (crude) indicator of the social capital of a community. Furthermore, once civic organizations are established for one purpose (for example, protesting against pollution), they can be flexibly adapted for a different purpose (responding to a disaster). Coleman (7) used the term "appropriable social organizations" to describe the phenomenon whereby an association established for one purpose could be later appropriated to serve a different purpose. In this manner, community organizations are more effective in "voicing" the demands of residents (1).

To summarize, in each of the processes described above—social contagion, informal social control, and collective efficacy—the individual benefits from their connection to a group (e.g., a network of parents in the community, or a network based on membership in a neighborhood association). However, social capital can also benefit people beyond the connected members of a network, that is, access to some resources in the group may be *nonexcludable*, which is a characteristic of what economists call a public good. For example, a resident may not belong to a local organization and yet benefit from the efforts of volunteers to clean up after a disaster; or an employee might be protected from catching the flu because other colleagues in the workplace conscientiously made appointments to get their shots at the clinic (herd immunity). That is, there are positive externalities (*spillover effects*) of social capital. It is this public goods aspect of social capital (i.e., the nonexcludability of consumption) that has attracted particular attention in contextual and multilevel studies of social capital and population health (see below).

THE DARK SIDE OF SOCIAL CAPITAL

Social capital, just like any form of capital, is agnostic with respect to whether the resource accessed through network connections is used for "good" or "bad" ends. Just as financial capital can be deployed for either good ends or bad ends—that is, for purchasing things that either promote health (dental floss) or damage health (cigarettes)—so social capital can cut both ways. Hence,

in the Framingham Study, smoking cessation was found to be contagious across social networks (17), but so was obesity (16); happiness is contagious (19), but so is depression (20). Advocates of social capital have sometimes been chastised for overlooking the duality—or "Janus-faced quality" (1)—of social capital (21). Sociability is a bit like motherhood and apple pie, and there is an unconscious bias to portray just its good side.

Portes (22) laid out some of the downsides of social capital in an influential essay. These include: exclusion of outsiders, excess claims made on group members, restrictions on individual freedoms, and downward-leveling norms. First, tightly knit and cohesive communities are often able to stay that way *because* they manage to keep outsiders from intruding. For example, Japanese society is frequently described as being highly cohesive (23). In turn, the social cohesion of the Japanese has been cited as a crucial ingredient of their longevity (24). But there is also a dark side to their solidarity that is hidden beneath the surface.

Cohesion in Japanese society can be traced back historically to the edicts of the Tokugawa *shogunate*, which maintained a continuous policy of isolationism for roughly two centuries (from 1633 to 1853) until the country was finally persuaded to open its doors by the naval bombardment of Commodore Matthew Perry. The Tokugawa rulers enforced seclusion over that period by threatening death to any foreigner caught on their soil. The legacy of this policy can be felt in the ethnic homogeneity of Japanese society, as well as the heavy dose of xenophobia that still lingers in the country's immigration policies to this day.[6] The dark side of social cohesion in Japanese society occasionally erupts into the open, such as happened in the shocking mass murder of five seniors, all in their seventies and eighties, in the remote village of Mitake in Yamaguchi Prefecture (in western Japan) in July 2013.[7] According to police reports, the 63-year-old perpetrator was a man who had returned to the village 20 years earlier to care for his aging parents. In confessing to his crimes, the man cited his motive as anger and resentment against the ostracism he suffered from his neighbors for failing to fit in. The Japanese even have a word for this type of social exclusion—*mura hachibu*. The term (literally meaning "village eight") refers to a custom dating back to feudal times when residents of rural communes in Japan united to help each other on ten specified occasions—such as weddings, illnesses, funerals, putting out fires, and so on. When anyone committed a major infraction, the punishment was to forbid communal assistance being extended to the offender on eight out of the ten events. Thus according to Prasol (25), "an 'eight' was a serious punishment at a time when rice-growing required mutual help; the outcasts were almost completely excluded from communal life, and could not survive very long" and "with time this communal tradition became widespread in Japanese society as a foundation for group behavior regulation."

Closer to home, the Boston busing crisis (1974–1988) was a series of riots that erupted in response to the court-ordered desegregation of public schools in traditionally Irish-American neighborhoods of the city of Boston, including South Boston, Charlestown, West Roxbury, Roslindale, and Hyde Park. That these riots were a manifestation of in-group solidarity in

6 For example, following the "Lehmann shock" of 2008, Japanese government offered thousands of dollars in cash to repatriate Brazilian immigrant workers (*New York Times*, April 22, 2009). These manufacturing workers—many of whom were Brazilian nationals *of Japanese descent*—were offered the cash so long as they promised never to return to Japan again.

7 http://en.wikipedia.org/wiki/Yamaguchi_arson_and_murders

response to a perceived external threat comes through in Michael Patrick MacDonald's searing autobiography of growing up in "Southie" (South Boston), "where everyone claimed to be Irish even if his name was Spinnoli . . . That's what we considered each other in Southie—family. There was always this feeling that we were protected, as if the whole neighborhood was watching our backs for threats, watching for all the enemies we could never really define. No 'outsiders' could mess with us" (26). In other words, the same moral resources such as trust and solidarity that keep cohesive groups together in the face of adversity can be used to exclude outsiders from gaining entry.

The second downside of social capital mentioned by Portes (22) is the excessive claims made on group members. According to Aldrich (1), social capital serves as a kind of informal insurance for members of a group; for example, enabling victims of a natural disaster to draw on preexisting support networks for financial, informational, and emotional assistance. The flip side of the ability of a member to draw down on network support during times of need is that someone else in the network is *supplying* those resources. And if the community is already resource-constrained, the exchange of favors can result in excessive levels of strain on group members who are constantly being called on to provide assistance to others, sometimes at high personal cost. Another illuminating example involving a double dark side scenario is the obligations placed on crime bosses to look out for the members of their fraternity. Crime organizations such as the mafia or the *yakuza* are clearly forms of social capital; they provide valuable resources to those who belong to them, even though they impose a negative externality on the rest of society. Yet even within this "dark" form of social capital, members are not immune from the downside of excessive obligations. In *Confessions of a Yakuza* (27), a biography of a Japanese crime boss, some of the numerous obligations of the criminal network are vividly described: supporting the welfare of relatives of *yakuza* who are sent away to prison, providing cash for funerals and medical costs, compensating rival gangs for misdemeanors committed by one's own underlings. To wit, "with all these obligations and looking after people and keeping up appearances, the boss of a *yakuza* gang was bound hand and foot, and however much money came in he never really had enough" (27).

The third type of downside mentioned by Portes (22) is the restriction on freedom that often coexists within a cohesive group. As explained earlier, informal social control helps to keep down deviance and antisocial behavior, but the dark side of this phenomenon is a community that is overcontrolling and intolerant of diversity. Lastly, the idea of "downward-leveling norms" refers to the phenomenon in which group cohesion pulls down outliers in the direction of the norm accepted by the group. This is perhaps most aptly expressed by the Japanese proverb "The nail that sticks out gets hammered down" (28). It is especially detrimental in the context of education, where a culture of deliberate underachievement can take root in already disadvantaged schools. This cultural outlook can be viewed as a protective mechanism exerted by the group to ensure that the student who strives too hard will not be disappointed by subsequent failure. If the prevailing norm in the school is that succeeding academically is "uncool," even capable students may end up seeking acceptance (i.e., conform to the norm) by following the example of their peers. Precisely this kind of "dark side" social capital is described by Jay MacLeod's classic ethnographic study of the Hallway Hangers (referring to the students who loitered in the hallways of their school instead of attending class) in *Ain't No Makin' It* (29).

BONDING, BRIDGING

An important distinction drawn by researchers is whether the social capital is of the bonding or bridging/linking type (30, 31). The distinction sometimes helps to explain why the dark side of social capital predominates over the good side in a given context. Bonding social capital refers to resources that are accessed within networks or groups in which the members share similar background characteristics such as class or race/ethnicity—that is, in network terminology, they are "homophilous." Bridging social capital by contrast refers to resources that are accessed across networks that cross (or "bridge") class, race/ethnicity, or other social characteristics.

The distinction helps to explain why some groups seem to have plenty of social capital but it does not help them to stay healthy. Thus, in many disadvantaged communities, strong bonding capital develops as an important survival mechanism for residents; but so long as the poor only have each other to turn to for support, they can remain "trapped." In Carol Stack's (32) classic ethnographic study of a poor African American community, the mutual exchange of support through kinship networks was identified as the primary mechanism for "getting by." The downside of this form of social insurance is that it extracts a high cost from community members, both financially as well as in terms of psychological strain. In one study in rural Birmingham, Alabama, Mitchell and LaGory (33) studied the mental health impact of social capital. They found that while high bonding capital—as measured by the level of trust and strength of ties between members of the community who shared similar race and socioeconomic backgrounds—was associated with *more* mental distress, the opposite was true for network ties to others who came from different race/class backgrounds (i.e., bridging capital).

These insights can help to explain the "inconsistent" findings that have been reported from time to time. For example, in a low-income neighborhood of Baltimore, children of mothers who reported lower levels of attachment to their community reported *fewer* behavioral and mental health problems (34), that is, it appeared to be paradoxically healthier not to have ties in the community. Similarly, in a study based in a working-class suburb in Adelaide, Australia, Ziersch and Baum (35) reported that greater involvement in community groups was associated with worse health. During qualitative interviews, the residents cited the strain of dealing with the daily problems that people brought to these group interactions. In addition to the excessive demands to help others, strong bonding capital often exhibits all of the other downsides enumerated by Portes (22), namely, *(1)* the down-leveling of norms, *(2)* the exercise of in-group solidarity to exclude members of out-groups, and *(3)* expectations of conformity and intolerance of diversity.

The emerging consensus from these studies is that bonding capital within disadvantaged settings can be as much of a liability as a force for the good. Consequently, it makes no sense to boost social capital in such settings without simultaneously investing heavily in other forms of investment—that is, economic and human capital development.

Bridging social capital, on the other hand, enables residents to access resources outside their immediate social milieu. It is a concept that explicitly links social capital to structural inequalities in power, resources, and authority. The two vignettes that opened this chapter both involved examples of bridging social capital. In the Kobe earthquake, strong bonding capital (in the form of *machizukuri* residents' associations) helped to deal with the immediate aftermath of the disaster (exchange of support between residents; recruiting volunteers). By contrast, during the prolonged

reconstruction phase, it was the establishment of new social capital—in the form of organizations that forged bridging ties between city authorities, nonprofit organizations, and community residents—that was most helpful in speeding the process of recovery. In the case of ethnic conflict in India, membership in the local branch of the Bharatiya Janata Party (BJP) boosts bonding capital among Hindus, while membership in the Muslim League does the same for Muslims. These types of social capital do not promote ethnic harmony. What makes the difference—according to Varshney (5)—is the presence of social capital that bridges the two groups.

Who benefits from bridging social capital can also vary according to context. In Japanese society, where formal institutions tend to be heavily dominated by men (for example, the country ranks 123rd among 189 countries in terms of female representation in political office), women might benefit more from bridging capital compared with men. In a survey of 4,000 residents of a midsized city in western Japan, Iwase and colleagues (36) inquired about participation in six different types of associations: Parents and Teachers Association, sports clubs, alumni associations, political campaign clubs, citizen's groups, and community associations. The authors distinguished between bonding and bridging social capital by asking participants about the homogeneity (with respect to gender, age, and occupation) of the groups to which they belonged. Bridging social capital (i.e., participation in groups involving people from a broad diversity of backgrounds) was strongly protective for self-rated health—but more strongly in women compared to men. Compared to women reporting zero participation, participation in groups with high bridging social capital was associated with a reduced odds of poor health (OR 0.25, 95% CI 0.11 to 0.55) after conditioning on demographic variables, socioeconomic status, and access to bonding capital. By contrast, bonding social capital was not consistently associated with better health in either sex (36).

MEASUREMENT OF SOCIAL CAPITAL

Conceptual approaches to the measurement of social capital can be encapsulated in a 2×2 matrix (Table 8.1). The rows of the matrix indicate the two distinct streams that have developed in the literature, distinguished by whether the researcher is emphasizing a *network-based* perspective on social capital or a *cohesion-based* perspective (10, 37). In the columns of the matrix, studies can be distinguished according to whether social capital is being analyzed at the individual level or the group level (Table 8.1).

Whether social capital ought to be considered an individual attribute or group attribute has been a topic of debate. Coleman (7) was quite explicit about his level of conceptual analysis: "Unlike other forms of capital, social capital inheres in the structure of relations between persons and among persons. It is lodged neither in individuals nor in physical implements of production" (p. 302). Contrasting with this stance, Portes (22) was equally prescriptive in calling for an exclusively individual level of analysis, identifying social capital as the resources that are derived from an individual's social network. From a population health perspective, we argue that there is utility of conceptualizing (and measuring) social capital at the group level. Harking back to the two vignettes at the start of the chapter, it is interesting to ask why some communities recover faster than others following exposure to a disaster and why some communities more effectively maintain peace and harmony. These are different questions than asking why some individuals

TABLE 8.1: Measurement approaches to assess social capital (from Kawachi, 2010)

	Level of Analysis	
Definition of social capital	Individual	Group
Network-based perspective	Position Generator Resource Generator	Whole social network analysis
Social cohesion-based perspective	Survey-based assessment of individual perceptions (e.g., trustworthiness of others) and behaviors (e.g., civic participation)	Survey responses aggregated to the group level.

fare better than others. At the end of the day, it is a forced dichotomy to argue exclusively for one or another level of analysis. Thus, we concur with Moore and colleagues (38) who opined that:

> "Debates on the appropriate level of analysis have often pitted communitarian against network approaches to social capital. Communitarian approaches have focused on social capital as the property of spatially-defined groups (e.g., neighborhoods, countries), whereas network approaches have tended to examine social capital at the personal or interpersonal levels. Nevertheless, as Bourdieu (6) emphasized, network social capital operates across both levels, since such capital is collectively owned but mobilized through individual and group actions. Hence, a network approach to social capital implies the consideration of how social capital operates across multiple levels of influence" (p. 192).

In the top left-hand cell of Table 8.1, we see examples of measurement approaches based on individual-level assessment of network-based capital—for example, Nan Lin's Position Generator (39), and van der Gaag and Snijder's Resource Generator (40). The Position Generator (39) is an instrument that assesses whether the individual is personally acquainted or connected to someone else who has an occupation that embodies valued resources represented by wealth, power, and prestige—for example, a doctor or lawyer. The assumption is that knowing someone doing these kinds of jobs correlates with the ability of the individual to access information and advice, instrumental support, or symbolic status. The Position Generator then yields additional indicators of an individual's social capital, such as "upper reachability," which references the highest level of prestige accessible to the individual in his/her network. Upper reachability is thus akin to the construct of "linking" social capital referred to earlier.

The Position Generator has proved most useful for studying how individuals can leverage their network connections to get ahead in society—for example, seeking the advice of a tax lawyer to deal with an IRS audit or receiving useful tips from a college admissions officer on how to prepare your child's application to a prestigious college. However, the implicit emphasis of the Position Generator on instrumental resources tends to limit its use in studying health outcomes. Thus, for example, the ability to access other individuals with prestigious jobs may not be relevant for understanding how network-mediated resources can promote health-related behaviors. In addition, the Position Generator is silent with regard to the resources provided by people who cannot be classified on a scale of occupational prestige, such as home-makers (41).

Yet another instrument to capture individual network-based social capital is van der Gaag and Snijder's Resource Generator (40). This instrument is characterized by a checklist approach in which the respondent is asked about the different types of skills or support that he/she can access via their network of friends and acquaintances—for example, whether the individual knows someone who can repair a broken-down car, or babysit their children in an emergency, or lend the person a large amount of money. In cross-sectional studies, the Resource Generator has been linked to health outcomes such as depressive symptoms (42) and self-rated health (43). There is some overlap between items found on the Resource Generator checklist with items that are found on existing *social support* instruments such as the Medical Outcomes Study Social Support Survey (44). However, an important distinction between them is that whereas the Resource Generator tends to focus on instrumental resources (information/advice, personal skills, money, and labor in kind), social support instruments give more emphasis on emotional support—for example, displays of affection, confiding in problems, relaxing together, and so forth. As discussed in a previous section, the Resource Generator seems to capture the types of support that an individual accesses via their weak ties, whereas Social Support instruments seem to tap into resources (such as emotional support) that people access via their intimate ties.

Turning to the upper right hand quadrant (on Table 8.1), the extension of the network-based approach to the group level is exemplified by whole social network analysis. In public health and epidemiology, the literature has been dominated by the assessment of ego-centered social networks (see Chapter 7). It is far less common to see analyses of whole social network data, and those that are publicly available seem to have been repeatedly data-mined to the point of diminishing returns, for example, the Framingham Study (45) and the National Longitudinal Survey of Adolescent Health (the Add Health Study) (46). The principal limitation of mapping the whole social network is the time and expense involved in interviewing all of the individuals (*alters*) nominated by each ego. While the boundaries of the network are readily identifiable in settings like schools or workplaces (or in a defined subpopulation such as an injection drug user's network), they are much less tractable in contexts such as a residential neighborhood. For example, in the Framingham Study, the researchers did not conduct a conventional whole social network assessment; instead they leveraged the fact that a high proportion of contact persons nominated by participants in the study (in case they were lost to follow-up) serendipitously happened to be also taking part in the same study within this tight-knit community in eastern Massachusetts (45). Bearing in mind these caveats, Lakon, Godette, and Hipp (47) offer examples of sociometric structural measures from whole network analysis that have potential relevance for the concept of social capital, including network-based analogs of social cohesion, bonding, and bridging social capital. What is clear from the description of these measures is that they rely on a complete and accurate mapping of all the ties that link the actors in a network. For example, a proposed structural sociometric measure of "cohesion" is the minimum number of actors who, if removed from the group, would disconnect the group. Figuring out this number could be quite exacting, since depending on the precise structure of the network, even the removal of one strategically located node could make a drastic difference to the connectivity of a group.

The lower row of the 2 × 2 matrix (Table 8.1) represents the approach to social capital measurement that has been most widely adopted (to date) in population health, that is, survey-based assessment of *social cohesion*. In surveys about social cohesion, the investigator makes no attempt to inquire about the respondent's social network connections. Instead the survey items inquire

about the potential availability of resources in the group, such as the reciprocal exchange of favors within the group to which the individual belongs. Generally speaking, surveys of social cohesion tap into two domains: (1) individual attitudes, perceptions, and cognitions about the group to which they belong, also referred to as *cognitive* social capital; and (2) actual behaviors (e.g., whether individuals participate in informal and formal social organizations), also referred to as *structural* social capital (48). The individual responses to survey items can then be analyzed at either the individual level (lower left-hand cell of the matrix on Table 8.1), or aggregated to the group (lower right hand cell) and analyzed as a property of the collective (e.g., the neighborhood or the workplace).

The social cohesion approach to measurement has been criticized for straying from the original network-grounded definition of social capital, that is, as "resources accessed through membership in networks" (49). This criticism has validity. However, given the formidable challenges of conducting whole social network assessment in contexts such as residential neighborhoods, we argue that it is a reasonable compromise to inquire instead about residents' perceptions of the availability of resources in the group, as well as perceptions of collective efficacy and informal social control. For example, in the Project on Human Development in Chicago Neighborhoods, a five-item "social cohesion" instrument was developed that asked respondents how strongly they agreed (on a five-point Likert scale) that "people around here are willing to help their neighbors," "this is a close-knit neighborhood," "people in this neighborhood can be trusted," "people in this neighborhood don't get along with each other," and "people in this neighborhood don't share the same values" (with the last two items being reverse-coded) (18). Individual responses to these items were then aggregated to the 343 Chicago neighborhoods, and then validated via the "ecometric" approach (50).[8]

The extent to which the individual items in a social cohesion scale (such as the Chicago survey instrument) actually overlap with network-based definitions of social capital is debatable. Sampson et al. (51) have argued that whereas network-based definitions of social capital tend to emphasize the resources obtained through private ties, the construct of collective efficacy (of which social cohesion constitutes a part, together with informal social control) taps into a neighborhood's ability to mobilize resources for the benefit of its members. Similarly, Carpiano (49), who advocates a network-based theory of neighborhood social capital, emphasizes four constituent types of resources: *(1)* the social support that residents can draw on from neighbors, *(2)* the leveraging of social connections to other residents in order to obtain useful information, *(3)* informal social control, and *(4)* residents' participation in neighborhood organizations. In short, there turns out to be a considerable degree of overlap in the constructs championed by the proponents of the social cohesion approach and the social network approach to measuring neighborhood social capital.

That said, one point of ongoing contention is whether *trust* ought to be considered a part of social capital. The question is whether trust can be considered as a moral resource inherent in social networks, or whether it is a predisposing factor (i.e., an antecedent) for social capital, not an integral part of it (48). Trust certainly lubricates the exchange of resources. For example, when an individual (let's call her Anna) approaches her friend Betty to ask for a cash loan, Betty

8 A number of other psychometrically validated instruments (based on the social cohesion approach) have been developed for use in field studies of social capital, and interested readers are referred to the review by Harpham (2008) (48).

will comply with the request if she trusts that Anna will pay her back. Moreover Betty's trust of Anna becomes enforceable if both Anna and Betty are friends with Christina, that is, there is network closure. If Anna then defaults on her loan, she will certainly have a hard time borrowing more money from Betty in the future, *but in addition* she further risks damaging her reputation with Christina, so that Christina is unlikely to agree to lending money to Anna as well. In this example, the network structure (the fact that Betty and Christina are friends of Anna, but they are also friends with each other) creates enforceable trust. That is, the members of this relationship triangle can trust other members to follow certain norms of behavior (i.e., repay cash loans from friends), and that trust is enforced by the threat of sanctions (i.e., ostracism) for violating the group's norms. George Homans noted as much in his 1958 classic on social behavior as exchange (52):

> When members of a group see another member as deviating, their interaction with him—communications addressed to getting him to change his behavior—goes up fast the more cohesive the group.
>
> If the deviant fails to change his behavior, the other members start to withhold social approval from him; the deviant gets low sociometric choice (translation: he becomes ostracized). But how plonking can we get? These findings are utterly in line with everyday experience. (52)

Trust also makes collective action more likely, that is, when members of a group can trust that others will not shirk their duty and free-ride on the effort of good citizens. In short, the stock of trust in a group is an intangible but crucial resource that enables resources to be mobilized and exchanged. To put it in stark terms, without trust, it is hard to imagine that any exchange of resources could take place between members of a network; the resources would remain locked up and frozen in the possession of individual actors.

The problem with trust arises when it is analyzed as an individual attribute (the lower left-hand cell of the 2 × 2 matrix in Table 8.1). One ambiguity that often goes unnoticed is the subtle but critical distinction between "trust of others" versus the "trustworthiness of others." Individual responses to questions such as: "In general, would you say that your neighbors can be trusted? (Strongly Agree…. Strongly Disagree)" cannot distinguish between an individual's tendency to trust others versus the actual trustworthiness of her neighbors. The former is a psychological trait (indeed the routine lack of trust of other people is called "cynical hostility," and has been shown to be a risk factor for poor health) (53). By contrast, social capital researchers are primarily interested in the question of whether being surrounded by trusting neighbors is good for their health, that is, whether a trustworthy environment lubricates the exchange of resources. Unfortunately, analyses based on individual perceptions of trust are open to both interpretations. For social capital research to add value beyond established findings in personality psychology, it needs to capture the trustworthiness of the group. One solution is to aggregate individual responses to the trust item to the level of the group (i.e., the lower right hand cell of Table 8.1), in which the group average value is assigned to each individual. By doing so, the averaged perception of trust is less likely to be influenced by individual variations in cynical hostility. We may argue that the resulting metric captures a collective characteristic of the group, for example, the neighborhood or the workplace.

A further criticism of the measurement of social cohesion via surveys is that the instruments frequently include elements that do not belong to the construct of social capital, but instead represent consequences of social capital, or remote proxies of social capital (48). For example, people's satisfaction with their neighborhoods strays far from the definition of social capital as "resources accessed via networks." Occasionally when survey data are lacking, researchers have resorted to the use of proxies such as crime statistics or voting participation. These are examples of the kind of "conceptual stretching" that Portes (22) warned about.

We conclude this section by briefly alluding to experimental approaches to measuring social capital. Economists—many of whom distrust survey responses to questions about trust—have advocated experimental approaches to measure constructs related to social capital, such as trust and cooperation. For example, Glaeser and colleagues (54) have proposed the "envelope drop" method to directly observe trusting behaviors. In this approach, stamped and addressed envelopes are dropped on random street corners in a neighborhood, and the experimenter can directly observe what proportion of the letters are picked up by anonymous strangers and mailed back to the addressee.

An alternative experimental approach is to study trusting and cooperative behavior in the context of "games" involving strategic interaction. For example, in the classic trust game, the first subject, A, is given a sum of money and offered the opportunity to pass some, all, or none of it to partner B. The experimenter promises to increase the transferred amount by some multiple before passing it on to B. Finally, B has the opportunity to return some, none, or all of the money back to A. In this experiment, the amount initially transferred by A is interpreted as a measure of trusting behavior. Other games, including the public goods game (a version of the prisoner's dilemma), are described by Anderson and Mellor (55). Reassuring to those who rely on surveys, there is evidence of convergent validity between survey-based and experimental measures of social capital; it has been found that individuals who self-report greater trust of others or higher participation in voluntary groups on surveys are also more likely to exhibit trusting and cooperative behaviors in experimental situations (56).

EMPIRICAL EVIDENCE

Empirical studies of social capital and health are too numerous to describe individually here. Instead we provide a summary of the main findings in this section. Interested readers are referred to systematic reviews that have been conducted on social capital and physical health (57), mental health (58), and health-related behaviors (59).

The first observation to make about the studies is that the majority have focused on neighborhood social capital as the "exposure," and that most of them have approached the measurement from a social cohesion perspective. Murayama et al. (60) have summarized the multilevel studies of neighborhood social capital and health based on a systematic review. The authors noted four pertinent findings with respect to the state of evidence:

- There is more consistent evidence linking individual perceptions about social cohesion in the neighborhood to health outcomes (lower left-hand cell of Table 8.1) compared with

evidence for a contextual effect of neighborhood cohesion on health (lower right hand cell). In multilevel studies, when level-1 perceptions are controlled for, the coefficient for the level-2 cohesion variable often becomes attenuated to statistical nonsignificance. For example, in a prospective study of 11,092 community-dwelling elderly individuals (65–84 years) in Shizuoka Prefecture, Japan, individual perception of community cohesion was associated with a reduced risk of all-cause mortality (HR = 0.78; 95% CI = 0.73, 0.84) as well as mortality from cardiovascular disease (HR = 0.75; 95% CI = 0.67, 0.84), pulmonary disease (HR = 0.66; 95% CI = 0.58, 0.75), and all other causes (HR = 0.76; 95% CI = 0.66, 0.89) (61). However, conditioning on individual perceptions of social cohesion, no statistically significant relationship was found between community cohesion (as a level-2 attribute) and mortality risk.

- The most consistent evidence across both individual and multilevel studies of social cohesion has been reported when self-rated health was examined as the outcome, that is, the single item measure asking individuals to rate their overall health as "Excellent, Very Good, Fair, or Poor." However, when self-rated health is used as the outcome, we end up with individual perceptions on both the left-hand and right-hand of regression equations, thereby raising the specter of common method bias, namely, the potential for confounding by unobserved individual characteristics, such as negative affectivity.
- The bulk of evidence to date has been based on cross-sectional designs, and there remains a dearth of prospective studies.
- Community social cohesion exhibits the "Janus-faced" characteristic described by Aldrich (1), that is, for some, social cohesion provides a health benefit, while for others it is either useless or even harmful. This implies a cross-level interaction between community cohesion and individual characteristics. For example, Subramanian, Kim, and Kawachi (62) found no overall association between community cohesion and mental health in the Social Capital Community Benchmark Survey. However, the test of cross-level interaction between community cohesion and individual trust was statistically significant. That is, for trusting individuals, living in a highly cohesive community appeared to be good for their mental well-being, but for mistrustful individuals, the opposite was true; they appeared to pay a penalty as a result of being surrounded by trusting neighbors.

Beyond the need for better-designed studies (e.g., longitudinal follow-up), the literature calls for greater specificity in the choice of health outcomes grounded in theory, as well as a more careful examination of treatment heterogeneity (i.e., for whom is social cohesion beneficial, and for whom is it harmful?). It would be fair to say that the first generation of studies on social capital relied on whatever secondary data happened to be conveniently available to the researcher. Consequently, the measurement of social capital tended to be based on proxies, and the selection of health outcomes not well matched to theory. For instance, the construct of informal social control was originally developed in the field of criminology to explain community variations in their ability to maintain social order. There is no immediately plausible reason why we would expect informal control to explain community variations in obesity.[9] That is not to deny that there may

9 For instance, we do not see the "obesity police" patrolling our neighborhoods—although one day it may come to that.

be *specific* health outcomes that are directly affected by informal control—for example, the ability of a community to stop public drinking by underage minors, or the ability of seniors to survive a heat wave.

In Klinenberg's (63) "social autopsy" of the 1995 heat wave in Chicago, one of the biggest risk factors for death among low-income seniors was remaining cooped up indoors and failing to seek relief at an emergency cooling station. It turned out that many seniors in disadvantaged communities were simply too afraid to come outdoors because of their fear of crime. Klinenberg contrasts the radically divergent mortality experiences of two adjoining neighborhoods in the west side of Chicago, where the decisive difference between them seems to have been the levels of informal social control predating the heat wave. In North Lawndale, residents felt unsafe and refused to come out to seek help, with the result that the heat-wave mortality rate reached 40 deaths per 100,000. In neighboring South Lawndale—where the death toll was one-tenth the rate of North Lawndale—Klinenberg's ethnographic interviews with residents revealed an active sense of informal control and collective efficacy. For example, a local priest interviewed by the author volunteered that "even though we have gangs, people still feel comfortable in the streets. You walk around and you see people sitting on the front steps everywhere" (63). Another local resident stated that: "We look out for each other in our neighborhood. If something is going on we'll see it and call each other or the police" (63). These sentiments are as close as we could get to lay expressions of the concept of collective efficacy and informal control. To summarize, even though informal social control was originally developed to explain community variations in crime, in specific circumstances it can explain health outcomes as well; we just need more specificity in the choice of health outcomes that match the theorized mechanisms.

THE SPATIAL DIMENSION OF SOCIAL CAPITAL

Besides the need to match the health outcome to the theoretically specified mechanism, researchers who study the neighborhood effects of social capital need to do a better job of incorporating the spatial dimension. Although it is standard practice to adopt administrative boundaries such as the Census tract or Census block group as the definition of a neighborhood, there is no reason to suppose that social interactions conform to such boundaries. Social interactions (which give rise to social capital) do not necessarily respect administrative boundaries—that is, there are likely to be spatial spillover effects. Failing to take account of this can induce the equivalent of exposure misclassification.

Using crime victimization as the outcome, Takagi et al. (64) contrasted two alternative approaches to analyzing the effects of neighborhood social capital in Tokyo City. In the first approach, the authors analyzed the data using conventional hierarchical regression analysis in which the level-2 boundaries were based on administrative definitions. In the alternative approach a spatial Durbin model was adopted, using an inverse-distance weighting matrix that assigned each respondent a unique level of "exposure" to social capital based on all other residents' perceptions.

In other words, the spatial approach attempts to address the problem of spatial spillovers by weighting the unique "force" of social capital felt by each individual according to the inverse of the distance between that individual and all other individuals living in the same locality, giving rise to an inverse-distance spatial-weighting matrix. In other words, the greater the distance between two individuals, the weaker the mutual "force" of social capital. Social capital in this study was assessed by survey responses to perceptions of generalized trust, reciprocity, and informal socializing with neighbors.

Using a survey conducted in one ward of Tokyo city, Takagi et al. (64) demonstrated that conditional on the individual's social network characteristics, residents in geographic locations with stronger social capital (as measured by the distance-weighted "force" of trust and norms of reciprocity) are protected from crime victimization. Strikingly, in the same dataset when the analysis is repeated using multilevel analysis (based on officially defined neighborhood boundaries), no association was found between community social capital and crime victimization. In other words, the study would have been reported as a null finding had it been based just on standard multilevel modeling.

In a second demonstration of the relevance of considering the spatial dimension of social capital, Takagi (65) constructed alternate buffer zones for each individual in a dataset, again disregarding administrative boundaries. To simplify the demonstration, Takagi used the perceived level of trust as the indicator of social capital (obtained from a mailed survey in one ward of Tokyo). The social influence of neighbors on each individual was then calculated as the average of trust expressed by all residents within different-sized circular buffer zones, which were calculated in 10-meter increments ranging from 50 m to 500 m. A piecewise regression analysis was then conducted for each buffer, which suggested that the association between trust and burglary victimization is *nonlinear* and U-shaped; that is, the protective effect of neighbors' trust on crime victimization was strongest at the most proximate distance (50 m), but waned for distances between 50 m and 499 m, before increasing again in strength beyond 500 m. What could account for this U-shaped association between distance and the effects of social capital?

Sociological theory about crime prevention provides a potential explanation. At the very intimate scale (<50 m), neighbors who live in close proximity to each other depend on mutual daily "management activities"—such as watching out for your neighbor's home when he is on vacation, and making sure that the mail and newspapers do not pile up in the driveway. But acts of reciprocity tend to fade with distance, that is, you might perform a favor for neighbors on either side of your home as well as the house opposite to yours on the other side of the street, but you'd be less likely to perform such acts for houses that are a block away. In contrast to this model of reciprocity on the intimate scale, social capital-related mechanisms such as collective efficacy arise at larger scales of social organization. In other words, it takes more than a couple of neighbors working together to mobilize collective action to solve shared problems. For example, mounting a neighborhood patrol or gathering petitions to lobby for more police protection are activities that require more than the voluntary efforts of a few concerned neighbors. Although the preceding examples use the case of crime, there is every reason to believe that a similar approach could be applied to health outcomes, and causal inference could be improved as a result of more explicit consideration of the spatial dimension of social capital.

WORKPLACE SOCIAL CAPITAL

A promising new direction for social capital research is represented by the extension of the concept to the workplace social environment (66). The workplace seems like a natural setting to examine the influence of social capital: it is a setting in which people are spending an increasing part of their daily lives, and it is a context in which many people form durable network ties. Indeed studies of workplace social capital have so far provided some of the most convincing empirical evidence linking social capital to health outcomes. These studies—particularly coming out of the Finnish Public Sector cohort—have been of high quality, featuring large sample sizes, prospective follow-up design, the use of validated and reliable social capital instruments, and linkage to validated medical records for ascertaining health endpoints. Three reports from this cohort merit particular attention.

Oksanen et al. (67) examined prospectively the association between workplace social capital and all-cause mortality during 5 years of follow-up among 28,043 workers employed in the Finnish public sector. Two waves of social capital surveys (in 2000–2002 and in 2004) were linked to national mortality registers through 2009. Social capital was measured through a validated eight-item social cohesion scale inquiring about norms of trust and reciprocity in the work unit as well as practices of collective action. In Cox proportional hazards models each point increase in the mean of repeated measurements of self-reported social capital (range 1–5) was associated with a 19% decrease in the risk of all-cause mortality (HR: 0.81, 95% CI: 0.66–0.99). The corresponding point estimate for coworker-assessed social capital was similarly protective (HR = 0.77, 95% CI: 0.50–1.20). The authors also took advantage of the repeated assessment of workplace social capital to perform a fixed effects analysis, that is, to analyze the impact of a change in workplace social capital on mortality risk controlling for all time-invariant observed and unobserved confounding characteristics. The fixed effects analysis yielded a point estimate for the odds ratio that was imprecise, but very consistent with the Cox regression estimate (OR = 0.81, 95% CI: 0.55–1.19).

In a separate analysis, Oksanen et al. (68) examined the association between social capital and incidence of hypertension (determined from record linkage to national health registers) among 11,777 male and 49,145 female employees who were free of hypertension at baseline. During a mean of 3.5 years of follow-up, male employees in work units characterized by low workplace social capital were 40–60% more likely to develop hypertension compared to men in work units with high levels of social capital. According to path analysis of the data adjusted for covariates, the association between low social capital and hypertension was partially mediated by higher risk of obesity (p-value for pathway = 0.02) and excess alcohol consumption (p = 0.03).

Not every health behavior or outcome in the Finnish Public Sector cohort turns out to be correlated with workplace social capital. For example, all-cause mortality (67) and poor self-rated health (69) are associated with work-unit social capital even after controlling for individual perceptions and other covariates. By contrast, new-onset depression (70) and smoking cessation (71) are not. Workplace social capital predicts the onset of new cases of hypertension (72) but does not predict medication adherence among individuals being treated for hypertension (68). There is thus much that still needs to be understood about the precise mechanisms linking workplace capital to specific health outcomes. Because of the dual-faced nature of social capital, it can

be difficult to predict whether its effects on workers' health will be beneficial or detrimental. For example, if friendships are formed while smokers step outside the office to get their nicotine fix, then social capital might not be helpful for smoking cessation. On the other hand, if a worksite smoking cessation intervention were introduced, smokers in a closely knit workplace could support each other to quit together. Much depends on an understanding of the local context.

A future challenge for research on the influence of the work environment is to understand the effects of the multiple social contexts in which the individual is embedded. Thus, a worker is simultaneously exposed to neighborhood contexts (where they live) and workplace contexts (where they work). In short, an association between workplace social capital and health could be confounded by the influence of people's neighborhood contexts. Alternatively, there may be cumulative influences—or even compensatory influences—of dual exposure to workplace and neighborhood environments. Disentangling these issues in future studies will require measurement of both contexts, as well as special analytical techniques (cross-classified multilevel analysis) to deal with this complexity.

ENDOGENEITY AND CAUSAL INFERENCE

In the years following the introduction of the concept of social capital to the field of population health (c. 1996), research has become more sophisticated with respect to both design and analysis. The first generation of research tended to be ecological in design (c. 1996–2000). The next generation of studies focused on individual-level as well as multilevel analysis (c. 2000–present time). The third generation of studies (since about 2007) has begun to address causal inference through methods such as instrumental variable estimation (15). Challenges of causal inference are common throughout social epidemiology—and indeed all of observational epidemiology—but they seem especially challenging for social behaviors, such as network formation and social participation. The reason is that most social behaviors—such as whether to trust someone sufficiently to lend him money, or whether to participate in a community organization—are rooted in the individual's choices and preferences, and hence endogenous in any equation linking social capital to health outcomes.

Overcoming the issue of endogeneity can be quite challenging. No amount of high-quality longitudinal data or statistical adjustment for covariates in a multilevel regression will suffice to convince a skeptic that endogeneity has been purged from the data. To give two simple examples, much of the network effects in the Framingham Study have come under attack because the analytical methods used by the researchers failed to take account of homophily, that is, the notion that people with shared characteristics tend to befriend each other (73, 74). Thus, when we observe a temporal association between two individuals who belong to the same network becoming obese, this might not be because of social contagion; it might be because "birds of a feather like to flock together" (perhaps the stigma of obesity leads overweight people to seek each other's company where they can feel more comfortable). A second example of potential endogeneity involves the frequently observed correlation between social participation and health. Once again, a temporal association does not prove that social participation promotes health. The two alternative explanations are that: *(1)* healthy people are more likely to join groups, and/or *(2)* the association

is confounded by unobserved heterogeneity, for example, temperament or personality or some other characteristic that acts as a common prior cause of social participation and later health.

One solution to cut this Gordian knot is to directly manipulate the exposure of interest, for instance, by randomizing the formation of friendship ties (such as happens in freshmen dormitories on some college campuses [75]), or launching a community-based program to encourage social participation on a cluster-randomized basis. Unfortunately, we do not always have the luxury of time (or funding) to conduct experiments; and hence, researchers have increasingly turned toward observing natural experiments as a way to step closer to causal inference. In the area of social capital, researchers have increasingly turned toward instrumental variable (IV) estimation.

Instrumental variable estimation has been long established in the field of economics and other social sciences, but it is a comparatively recent import into the field of social epidemiology (76). The principle behind IV estimation is to find variables "in nature" that cause variation in the level of the exposure of interest (in our instance, some indicator of social capital). The two requirements of a valid instrumental variable are that: *(1)* it should be correlated with the exposure to a degree that it captures sufficient variation in the treatment; and *(2)* it must have no direct effect on the outcome (the so-called exclusion restriction) (77). For a more detailed explanation of the IV approach, the reader is referred to Chapter 2, which discusses the use of state compulsory schooling laws as an instrument to identify the causal effects of education on health outcomes.

A range of instruments has been tried in the area of social capital, with varying degrees of plausibility and persuasive appeal (for a detailed summary, see Kawachi et al. [15]). For example, duration of residence in the community has been used to instrument perceptions of trust (78). The argument here is that the more stable the duration of residence in a place, the more opportunities an individual will have to interact with neighbors and to form trusting relationships. In order for the instrument to be valid, there must not be any relationship between duration of residence and the health outcome (in this case, self-rated health), *other than that passing through* trust. Other researchers have used indicators of population heterogeneity—such as religious fractionalism—as instruments for social cohesion (79–81). Again, whether religious fractionalism succeeds as an instrument depends on the argument that the effect of fractionalism on health is completely mediated by social cohesion, and that there are no direct paths from religious fractionalism to health (the exclusion restriction). In other research involving a sample of community-dwelling older adults in Argentina, Ronconi et al. (82) used access to local transport to instrument the level of informal socializing among seniors.

Ichida et al. (83) sought to examine whether participation in local community centers improved the health of the elderly. In this study, based in one municipality in Japan, the civic authorities decided to establish half a dozen community centers in the neighborhood to encourage seniors to socialize with each other. Even though the study design was longitudinal (i.e., data were available both prior to the opening of the community centers as well as afterward), the researchers reasoned that the probability of social participation would be endogenous; in other words, that healthy and sociable people would selectively participate. In order to get around this problem the authors used *distance* to the nearest community center as an instrument for social participation. The argument is that if the individual happened to live close to a newly opened community center, he or she would be more likely to participate (because of the convenience of getting there); and that whether someone lived close or far away from a community center ought not to have any direct relationship to health (except through variation in the extent of social participation).

As Ichida et al. (83) found, there is a strong correlation between distance and participation—if a community center opened nearby, the seniors in the study were more likely to go there to socialize with their neighbors. This instrument would not work if the town authorities deliberately selected the location of the centers based on the strength of lobbying by local residents. But this was not the case. The community centers were established primarily based on convenience, that is, wherever underutilized space happened to be available, such as in day-care centers after hours. The authors therefore argued that the distance of each resident to the nearest community center was more or less random.

Once an instrument is identified, the estimation procedure proceeds in two stages. In the first stage of the regression, the endogenous exposure is regressed on its predictors (i.e., "instrumented"). In the second stage, the outcome (health) is regressed on the instrumented values of the exposure, controlling for additional observed covariates. In the example by Ichida et al. (83), the IV analysis suggested a strong association between social participation and health; the odds ratio of excellent or good self-rated health among people who participated in the community centers was 2.52 (95% CI: 2.27 to 2.79).

Because most IV analyses have thus far been limited to examining the effects of individual-level social capital—for example, individual variations in civic participation or individual perceptions of trust—there remains a significant gap in the literature. In other words, the target of inference in most studies has been the individual. There is a dearth of studies that have sought to identify the causal effect of *contextual*-level social capital on health outcomes. This approach presents a formidable challenge, since IV estimation within a multilevel analytical framework requires two sets of endogenous treatments—one at the individual level and the other at the group level.

SOCIAL CAPITAL INTERVENTIONS

The ultimate proof of the utility of the concept of social capital must come from demonstrations that observational evidence can be translated into effective interventions to improve health outcomes. Spencer Moore and colleagues (38) have laid out a useful typology to describe different kinds of interventions that have targeted social capital. In the first type, the aim of the intervention is to *build new forms* of social capital (such as building brand-new community centers, described in the previous section). In the second type, social capital is the *channel* (i.e., mediating variable) through which another, unrelated intervention is predicted to affect health. For example, many kinds of microcredit and microfinance interventions have been introduced in resource-poor settings to stimulate economic development. Depending on the way in which the microfinance is set up, a *by-product* of the intervention can be a strengthening of social capital (84). In the third type of intervention, social capital is treated as the *segmenting* device (i.e., a moderating variable), to predict the success or failure of other community-based interventions.

A growing number of interventions have sought to boost social capital directly by creating new network ties and strengthening social interactions in the community. For example the Experience Corps was a community-based intervention based in Baltimore, Maryland, that sought to mobilize retirees to volunteer as teachers' assistants in public elementary schools (85). The program

was couched as an attempt to build new network connections that bridged generations (between seniors and schoolchildren) as well as teachers, parents, and volunteers (86, 87). The evaluations of this program suggested a "win/win" result, that is, the program succeeded in elevating the senior volunteers' level of physical activity and functional mobility as well as the children's academic scores. A program closely modeled on the Experience Corps was subsequently rolled out in Japan—called the REPRINTS program—in which retired seniors were invited to volunteer in kindergartens and schools as teachers' assistants. Extensive evaluation of the REPRINTS program—described by Murayama et al. (88)—showed positive spillover benefits beyond the seniors and the schoolchildren, that is, teachers as well as the parents of children became more engaged in the education of children. In short, programs such as the Experience Corps and REPRINTS suggest a viable form of intervention to promote productive aging through fostering social connections at the same time as leveraging the human capital embodied in the growing segment of the "graying" population.

In an altogether different and more challenging social context, Brune and Bossert (89) conducted a 2-year intervention to build social capital in three postconflict communities in Nicaragua. In the aftermath of that country's extended civil war (1981–1989), many communities, especially in rural areas, have been riven by distrust and violence, not least because of the resettlement of former pro-Sandinista forces in close proximity to former opponents of the Sandinista regime. In this challenging context, Brune and Bossert (89) implemented in two communities an intervention that was designed to strengthen social cohesion. The intervention included components that: *(1)* developed management and leadership capacities in the villages, with the goal of strengthening community organization and self-management; *(2)* encouraged the development of higher levels of household participation in community activities; and *(3)* increased trust among community residents as well as between the community and local public institutions. According to the authors,

> While the interventions were dynamic and tailored to the specific needs and contexts of individual communities, they had to meet the following broad requirements: (1) build on existing organizations in the community rather than impose new organizations; (2) develop participation mechanisms that encourage increased and continuing attendance at meetings and encourage broad participation in project activities; (3) develop communication, consensus building, and conflict resolution skills both in the community organization and within the wider community to build higher levels of trust within the community; (4) encourage decision-making and empowerment of community members, especially those who have not participated previously; and *(5)* create enduring ties of support with organizations outside the community. (89)

Two years after the intervention, compared with the control community (which received no training), the intervention communities reported elevated perceptions of social cohesion (the belief that neighbors were ready to assist in times of need), increased likelihood of working together with neighbors on projects to benefit the community, and increased likelihood of respondents contacting local health officials about a problem in their neighborhood. Following the intervention, the researchers also found that higher levels of social capital were significantly associated with some positive health behaviors. The behavioral/structural components of social capital (including

participation in groups and social networks) were associated with more desirable individual health behaviors such as the use of modern medicine to treat children's respiratory illnesses. Attitudinal components of social capital were positively linked to community health behaviors such as working on community sanitation campaigns (89).

In the second type of social capital intervention suggested by Moore and colleagues (38), building social capital is not the direct target of the intervention, but rather it is an anticipated by-product of another intervention. For example, when urban planners improve the quality of recreational spaces in a city in order to promote physical activity among residents, an anticipated benefit on the side is an increase in social interactions. Another example is microfinance programs, which are typically used to stimulate the economic development of poor communities. In other words, microfinance primarily seeks to improve economic outcomes, but the lending programs are often also bundled with other community-based social interventions. Pronyk and colleagues (90, 91) conducted a cluster-randomized trial in rural South Africa that combined group-based microfinance with participatory gender and HIV prevention training with the goal of bolstering community solidarity. After two years of the intervention, the researchers found increases in both the cognitive and structural dimensions of social capital (measured by intensity of participation in community organizations and perceived levels of reciprocity, solidarity, and collective action). In turn, increased levels of cognitive social capital were found to be associated with higher condom use and lower HIV prevalence among men and women. Increased structural social capital (civic participation) was associated with a protective trend in risk behavior; however, in the same program, higher participation was also associated with increased rates of HIV infection. That is, the program produced some unanticipated side effects as well. As we noted in our discussion of the dark side of social capital, interventionists need to be cognizant of its dual nature, and mindful of the challenges of balancing the beneficial effects of social capital with its possibly harmful side effects.

The third and last use of social capital in an intervention is to treat the level of social capital as a segmenting (moderating) variable that can influence the comparative success (or failure) of other community-based interventions. For example, in the area of disaster research, it is broadly acknowledged that there is wide variability in the recovery and resilience of affected communities (1). Some of this variability can be explained by variations in the stocks of community social capital predating exposure to disaster (1). Accordingly, a social capital inventory should routinely be incorporated into disaster preparedness planning and needs assessment of communities that are vulnerable to disaster (4).

SOCIAL CAPITAL AND SOCIAL POLICY

Whenever a new policy idea comes along it seems prudent to temper uncritical enthusiasm with a degree of skepticism. It is incumbent on the researcher to be vigilant against the "capture" of new ideas by politicians who seek to push a hidden agenda. This explains the critical reaction that social capital received when it was adopted by the World Bank (as well as by "Third Way" politicians) in the 1990s. The language of "capital" attracted neoliberals with an ideological bent toward giving the market a greater role in social policy, while the "social" part attracted those who wished

to see a greater role for communitarian ideals. The list of criticisms directed toward social capital as a tool for social policy includes the following:

- *It seems to monetize social relationships*. Money can't buy love (to quote the Beatles), and the value of social relationships ought to be kept out of the realm of economic transactions. Appending the word "capital" to "social" represents an unwarranted intrusion of market ideas into the social world. To quote the vocal critic of social capital, Ben Fine (92): "social capital is a form of peripheral colonisation, incorporating all social theory other than economics. Whilst presenting itself as the opposition to economics imperialism, it offers feeble resistance because it has no alternative. Indeed, it prepares social theory for the colonising advance of the economic approach" (p. 799).[10]
- *It is being used as an excuse to cut costs* (21). Social capitalists are accused of arguing along the lines of: "If only communities could just help each other, there would be less need for welfare spending and other kinds of public assistance." Indeed according to McKnight (93), welfare programs are even to *blame* for "crowding out" social cohesion. In *The Careless Society: Community and Its Counterfeits* (93), McKnight argued that government provision of welfare services saps our duty to care for each other and erodes societal norms of mutual assistance, voluntarism, and community competence. In short, the welfare state fosters a nation of "clients," commodifying the kinds of support that members of communities used to provide for each other. Contrary to these claims, empirical analyses tend to show the opposite tendency, that is, strong welfare regimes boost social cohesion. Thus, Rostila (94) demonstrates that in the EU region, the more that a country spends in the aggregate on social protection, the higher are the levels of informal social participation and membership in civic associations, and the higher the level of social trust.
- *It ignores structural inequalities in power*. All this talk of horizontal bonding is nice, but so long as we ignore fundamental questions like "Who even gets to associate with whom?" the discourse on social capital will, at best, end up distracting policymakers' attention from focusing on inequalities, and at worst result in blaming the victims (95). In a word, social capital does not sprout in a vacuum or rain down from the sky like random cloudbursts; rather, social capital is itself shaped by broader structural forces operating at the level of society, such as historical patterns of residential mobility (e.g., the influx of immigrants, shifts in local labor markets), municipal investment in housing and local infrastructure, as well as policies that perpetuate residential segregation or the planned shrinkage of services and amenities (96). Ignoring these structural dimensions could condemn communities to remain trapped in disadvantage, no matter how much effort they invest in strengthening community bonds.
- *It is being oversold as a panacea for public health problems*. Enthusiastic advocates ignore or downplay the dark side of social capital. Strengthening social cohesion could backfire and just as easily result in more incidents of intolerance against outsiders (or even insiders who don't conform to the community norm).

10 In Ben Fine's amusing ditty, with his apologies to Philip Larkin: "They fuck you up with social cap. They may not mean to but they do. They fill you up with faults on tap/And add some extra, just for you."

- *It reflects middle-class values, and outdated ones at that.* Critics attribute the popularity of social capital to a vaguely articulated yearning for the values of the past, a return to some idealized notion of "community" that tugs at our mystic chords of memory. The problem is that it all depends on *whose* vision of the past we are talking about. Many of us may not wish to return to the "good old days" of civic boosters and enforced conformity of the type satirized by Sinclair Lewis in his novels like *Babbit* and *Main Street.* Although the discourse on social capital has been attacked for expressing fundamentally middle-class (and majority "white") values (97), it would be a mistake to dismiss the concept on those grounds alone. As Gilbert and Dean (98) have argued, the scholarship on social capital can be enriched by considering how African American communities have wielded collective efficacy over history to combat discrimination and oppression. Rather than calling for an abandonment of the concept, the scholarship on social capital and health needs to acknowledge and incorporate the dimension of race/ethnicity in order to better understand the role of community organization and political advocacy within black communities.

By any measure, this is a formidable list of criticisms that challenge the utility of social capital as a social policy tool. Nevertheless, a few principles and lessons can be extracted from the debate that might still prove to be useful for guiding future policy (assuming that we are prepared to avoid tossing the baby out with the bathwater). First, social capital cannot be translated into simplistic prescriptions such as exhorting community members to pull themselves together. Intervening on social capital must be thought of as a *complement* to broader structural interventions (such as improving access to local labor markets), not as a replacement for them (31). Careful historical analysis—such as Szreter and Woolcock's (31) discussion of the role of social capital in shaping the sanitary reforms in nineteenth-century Britain—show how politics and power relations can be brought back into the analysis of social capital and health. Secondly, there is no magic bullet or standard recipe book for how to intervene. Any policy needs to carefully consider the local context, as well as history. There is unlikely to be a "one-size-fits-all" prescription for strengthening social capital. Different *types* of social capital will likely matter for different objectives. For example, widely scattered weak ties are more effective at disseminating information, whereas strong and dense connections are more effective for collective action (99). As Sobel (100) cautions: "People apply the notion of social capital to both types of situation. Knowing what types of networks are best for generating social capital requires that one be specific about what the social capital is going to be used to do." Thus theory would suggest that it is not sufficient—and is perhaps even harmful—to strengthen bonding social capital among unemployed youth. A more helpful policy would seek to build bridging capital between unemployed youth and employed adults to provide access to role models and mentoring (101).

Any proposal to build social capital from scratch needs to pay close attention to the distribution of costs and benefits, including the possibility of unintended consequences. Mobilizing the exchange of social network resources always implies that *someone* in the network is being called on to provide those resources. If the network is already deprived with respect to the availability of resources, exhorting members to "give more" will only generate more strain and frustration. A gendered analysis would also lead us to predict that the obligation to provide support will fall disproportionately on the shoulders of women. Lastly, and most importantly, a social capital

investment strategy requires more than the donated efforts of volunteers. A sustainable strategy often requires joint investment from government, the nonprofit sector, and the private sector. Social capital cannot be seen as a cheap alternative to government spending. It costs money to support local organizations, to invest in human capital (e.g., the training and development of community leaders; paying for volunteers), and to build local infrastructure.

CONCLUSION

In this chapter we have provided an interim summary of the research on social capital as a social determinant of health. It is still a work in progress, based on the comparative recency of the concept in the field of population health. We have highlighted some potentially promising directions for further research, including: *(1)* the need to strengthen causal inference by leveraging quasi-experimental designs; *(2)* the extension of the concept to study health promotion in diverse contexts such as the workplace; *(3)* the need for more studies based on network-based approaches to measure social capital; and *(4)* the call for interventions that can demonstrate the utility of the concept for health improvement, taking into full consideration the "dark side" of social capital.

REFERENCES

1. Aldrich DP. Building resilience: social capital in post-disaster recovery. Chicago: University of Chicago Press; 2012.
2. Nakagawa Y, Shaw R. Social capital: a missing link to disaster recovery. Int J Mass Emerg Disasters. 2004;22(1):5–34.
3. Kawachi I, Subramanian S. Measuring and modeling the social and geographic context of trauma. J Traumatic Stress. 2006;19(2):195–203.
4. Koh H, Cadigan R. Disaster preparedness and social capital. In: Kawachi I, Subramanian S, Kim D, editors. Social capital and health. New York: Springer; 2008. pp. 273–85.
5. Varshney A. Ethnic conflict and civic life: Hindus and Muslims in India. New Haven: Yale University Press; 2002.
6. Bourdieu P. The forms of capital. In: Richardson J, editor. The handbook of theory: research for the sociology of education. New York: Greenwood Press; 1986. pp. 241–58.
7. Coleman JS. Foundations of social theory. Cambridge, MA: Harvard University Press; 1990.
8. Loury G. The economics of discrimination: getting to the core of the problem. J Am Public Policy. 1992;1:91–101.
9. Putnam RD. Bowling alone: the collapse and revival of American community. New York: Simon and Schuster; 2000.
10. Kawachi I. Social capital and health. In: Bird C, Fremont A, Zimmermans S, Conrad P, editors. Handbook of medical sociology. 6th ed. Nashville, TN: Vanderbilt University Press; 2010. pp. 18–32.
11. Bannock G, Baxter R, Rees R. The Penguin dictionary of economics. Harmondsworth, England: Penguin Books; 1972.

12. Arrow KJ. Observations on social capital. In: Dasgupta S, editor. Social capital: a multifaceted perspective. Washington, DC: World Bank; 1999. pp. 3–5.
13. Moore S, Bockenholt U, Daniel M, K F, Kestens Y, Richard L. Social capital and core neighborhood ties: a validation study of individual-level social capital measures of neighborhood social connections. Health and Place. 2011;17:536–44.
14. Legh-Jones H, Moore S. Network social capital, social participation, and physical inactivity in an urban adult population. Soc Sci Med. 2012;74:1362–7.
15. Kawachi I, Takao S, Subramanian S. Global perspectives on social capital. New York: Springer; 2013.
16. Christakis N, Fowler J. The spread of obesity in a large social network over 32 years. N Engl J Med. 2007;357(4):370–9.
17. Christakis N, Fowler J. The collective dynamics of smoking in a large social network. N Engl J Med. 2008;358(21):2249–58.
18. Sampson R, Raudenbush S, Earls F. Neighborhoods and violent crime: a multilevel study of collective efficacy. Science. 1997;64:918–24.
19. Fowler J, Christakis N. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. BMJ. 2011;337:a2338.
20. Rosenquist J, Fowler J, Christakis N. Social network determinants of depression. Mol Psychiatr. 2011;16(3):273–81.
21. Pearce N, Smith GD. Is social capital the key to inequalities in health? Am J Public Health. 2003;93(1):122–9.
22. Portes A. Social capital: its origins and application in modern sociology. Annu Rev Sociol. 1998;24:1–24.
23. Takao S. Research on social capital and health in Japan: a commentary on Ichida and on Fujisawa. Soc Sci Med. 2009;69(4):509–11.
24. Marmot M, Smith G. Why are the Japanese living longer? BMJ. 1989;299(6715):1547–51.
25. Prasol A. Modern Japan: origins of the mind—Japanese traditions and approaches to contemporary life. Singapore: World Scientific Publishing; 2010.
26. MacDonald MP. All souls: a family story from Southie. Boston, MA: Beacon Press; 1999.
27. Saga J. Confessions of a yakuza. Tokyo: Kodansha International; 1989.
28. De Mente BL. Japan's cultural code words. Tokyo: Tuttle Publishing; 2004.
29. McLeod J. Ain't no makin' it. Boulder, CO: Westview Press; 2004.
30. Gittell R, Vidal R. Community organizing: building social capital as a development strategy. Thousand Oaks, CA: Sage Books; 1998.
31. Szreter S, Woolcock M. Health by association? Social capital, social theory, and the political economy of public health. Int J Epidemiol. 2004;33(4):650–67.
32. Stack C. All our kin: strategies for survival in a black community. New York: Harper & Row; 1974.
33. Mitchell C, LaGory M. Social capital and mental distress in an impoverished community. City and Community. 2002;1:195–215.
34. Caughy M, O'Campo P, Muntaner C. When being alone might be better: neighborhood poverty, social capital, and child mental health. Soc Sci Med. 2003;57:227–37.
35. Ziersch A, Baum F. Involvement in civil society groups: is it good for your health? J Epidemiol Comm Health. 2004;58:493–500.
36. Iwase T, Suzuki E, Fujiwara T, Takao S, Doi H, Kawachi I. Do bonding and bridging social capital have differential effects on self-rated health? A community based study in Japan. J Epidemiol Community Health. 2012;66(6):557–62.
37. Kawachi I, Wamala S. Commentary: social capital and health—making the connections one step at a time. Int J Epidemiol. 2006;35(4):989–93.

38. Moore S, Salsberg J, Leroux J. Advancing social capital interventions from a network and population health perspective. In: Kawachi I, Takao S, Subramanian S, editors. Global perspectives on social capital and health. New York: Springer; 2013. pp. 189–203.
39. Lin N. Social capital: theory and research. New York: Aldine de Gruyter; 2001.
40. Van der Gaag M, Snijders T. The Resource Generator: measurement of individual social capital with concrete items. Soc Networks. 2005;27:1–29.
41. Van der Gaag M, Webber M. Measurement of individual social capital: questions, instruments, and measures. In: Kawachi I, Subramanian S, Kim D, editors. Social capital and health. New York: Springer; 2008. pp. 29–49.
42. Webber M, Huxley P. Measuring access to social capital: the validity and reliability of the Resource Generator-UK and its association with common mental disorder. Soc Sci Med. 2007;65(3):481–92.
43. Kobayashi T, Kawachi I, Iwase T, Suzuki E, Takao S. Individual-level social capital and self-rated health in Japan: an application of the Resource Generator. Soc Sci Med. 2013;85:32–7.
44. Sherbourne C, Stewart A. The MOS social support survey. Soc Sci Med. 1991;32(6):705–14.
45. Christakis N, Fowler J. Connected: the surprising power of our social networks and how they shape our lives. New York: Little Brown; 2009.
46. Bearman P, Moody J. Suicide and friendships among American adolescents. Am J Public Health. 2004;94(1):89–95.
47. Lakon C, Godette D, Hipp J. Network-based approaches for measuring social capital. In: Kawachi I, Subramanian S, Kim D, editors. Social capital and health. New York: Springer; 2008. pp. 63–81.
48. Harpham T. The measurement of community social capital through surveys. In: Kawachi I, Subramanian S, Kim D, editors. Social capital and health. New York: Springer; 2008. pp. 51–62.
49. Carpiano RM. Actual or potential neighborhood resources for health: what can Bourdieu offer for understanding mechanisms linking social capital to health? In: Kawachi I, Subramanian S, Kim D, editors. Social capital and health. New York: Springer; 2008. pp. 83–93.
50. Raudenbush S. The quantitative assessment of neighborhood social environments. In: Kawachi I, LF B, editors. Neighborhoods and health. New York: Oxford University Press; 2003. pp. 112–31.
51. Sampson R, Raudenbush S, Earls F. Beyond social capital: spatial dynamics of collective efficacy for children. Am Sociol Rev. 1999;64:633–60.
52. Homans G. Social behavior as exchange. Am J Sociol. 1958;63(6):597–606.
53. Barefoot J, Larsen S, Von Der Lieth L, Schroll M. Hostility, incidence of acute myocardial infarction, and mortality in a sample of older Danish men and women. Am J Epidemiol. 1995;142(5):477–84.
54. Glaeser E, Laibson D, Scheinkman J, Soutter C. Measuring trust. Q J Econ. 2000;115(3):811–46.
55. Anderson L, Mellor J. The economic approach to cooperation and trust: lessons for the study of social capital and health. In: Kawachi I, Subramanian S, Kim D, editors. Social capital and health. New York: Springer; 2008.
56. Anderson L, Mellor J, Milyo J. Social capital and contributions in a public goods experiment. Am Econ Rev. 2004;94:373–76.
57. Kim D, Subramanian S, Kawachi I. Social capital and physical health: a systematic review of the literature. In: Kawachi I, Subramanian S, Kim D, editors. Social capital and health. New York: Springer; 2008. pp. 139–90.
58. Almedom A, Glandon D. Social capital and mental health: an updated interdisciplinary review of primary evidence. In: Kawachi I, Subramanian S, Kim D, editors. Social capital and health. New York: Springer; 2008. pp. 191–214.
59. Lindstrom M. Social capital and health-related behaviors. In: Kawachi I, Subramanian S, Kim D, editors. Social capital and health. New York: Springer; 2008. pp. 215–38.

60. Murayama H, Fujiwara Y, Kawachi I. Social capital and health: a review of prospective multilevel studies. J Epidemiol. 2012;22(3):179–87.

61. Inoue S, Yorifuji T, Takao S, Doi H, Kawachi I. Social cohesion and mortality: a survival analysis of older adults in Japan. Am J Public Health. 2013;103(12):e60–6.

62. Subramanian S, Kim D, Kawachi I. Social trust and self-rated health in US communities: multilevel analysis. J Urban Health. 2002;79(4 Suppl 1):S21–34.

63. Klinenberg E. Heat wave: a social autopsy of disaster in Chicago. Chicago: Chicago University Press; 2002.

64. Takagi D, Ikeda K, Kawachi I. Neighborhood social capital and crime victimization: comparison of spatial regression analysis and hierarchical regression analysis. Soc Sci Med. 2012;75(10):1895–902.

65. Takagi D. Neighborhood social capital and crime In: Kawachi I, Takao S, Subramanian S, editors. Global perspectives on social capital. New York: Springer; 2013. pp. 143–66.

66. Oksanen T, Suzuki E, Takao S, Vahtera J. Workplace social capital and health. In: Kawachi I, Takao S, Subramanian S, editors. Global perspectives on social capital and health. New York: Springer; 2013. pp. 23–63.

67. Oksanen T, Kivimäki M, Kawachi I, Subramanian S, Takao S, Suzuki E, et al. Workplace social capital and all-cause mortality: a prospective cohort study of 28,043 public-sector employees in Finland. Am J Public Health. 2011;101:1742–8.

68. Oksanen T, Kawachi I, Kouvonen A, Suzuki E, Takao S, Sjosten N, et al. Workplace social capital and adherence to antihypertensive medication: a cohort study. PLoS One. 2011;6(9):e24732.

69. Oksanen T, Kouvonen A, Kivimäki M, Pentti J, Virtanen M, Linna A, et al. Social capital at work as a predictor of employee health: multilevel evidence from work units in Finland. Soc Sci Med. 2008;66:637–49.

70. Kouvonen A, Oksanen T, Vahtera J, Stafford M, Wilkinson R, Schneider J, et al. Low workplace social capital as a predictor of depression: the Finnish Public Sector Study. Am J Epidemiol. 2008;167:1143–51.

71. Kouvonen A, Oksanen T, Vahtera J, Väänänen A, De Vogli R, Elovainio M, et al. Work-place social capital and smoking cessation: the Finnish Public Sector Study. Addiction. 2008;103:1857–65.

72. Oksanen T, Kawachi I, Jokela M, Kouvonen A, Suzuki E, Takao S, et al. Workplace social capital and risk of chronic and severe hypertension: a cohort study. J Hypertens. 2012;30(6):1129–36.

73. Cohen-Cole E, Fletcher J. Detecting implausible social network effects in acne, height, and headaches: longitudinal analysis. BMJ. 2008;337:a2533.

74. Lyons R. The spread of evidence-poor medicine via flawed social-network analysis. Statistics, Politics, and Policy. 2011;2(1).

75. Yakusheva O, Kapinos K, Weiss M. Peer effects of the freshman 15: evidence from a natural experiment. Econ Hum Biol. 2011;9:119–32.

76. Glymour M. Natural experiments and instrumental variable analysis in social epidemiology. In: Oakes JM, Kaufman JS, editors. Methods in social epidemiology. San Francisco, CA: John Wiley & Sons; 2006. pp. 429–60.

77. Angrist J, Pischke J. Mostly harmless econometrics. Princeton, NJ: Princeton University Press; 2009.

78. Schultz J, O'Brien A, Tadesse B. Social capital and self-rated health: Results from the US 2006 social capital survey of one community. Soc Sci Med. 2008;67:606–17.

79. D'Hombres B, Rocco L, Suhrcke M, McKee M. Does social capital determine health? Evidence from eight transition countries. Health Econ. 2010;19:56–74.

80. D'Hombres B, Rocco L, Suhrcke M, Haerpfer C, McKee M. The influence of social capital on health in eight former Soviet countries: why does it differ? J Epidemiol Comm Health. 2011;65:44–50.

81. Kim D, Baum C, Ganz M, Subramanian S, Kawachi I. The contextual effects of social capital on health: A cross-national instrumental variable analysis. Soc Sci Med. 2011;73:1689–97.

82. Ronconi L, Brown T, Scheffler R. Social capital and self-rated health in Argentina. Health Econ. 2012;21:201–8.

83. Ichida Y, Hirai H, Kondo K, Kawachi I, Takeda T, Endo H. Does social participation improve self-rated health in the older population? A quasi-experimental intervention study. Soc Sci Med. 2013;94:83–90.
84. Kondo N, Shirai K. Microfinance and health. In: Kawachi I, Takao S, Subramanian S, editors. Global perspectives on social capital and health. New York: Springer; 2013. pp. 239–75.
85. Fried L, Carlson M, Freedman M, Frick K, Glass T, Hill J, et al. A social model for health promotion for an aging population: initial evidence on the Experience Corps model. J Urban Health. 2004;81:64–78.
86. Glass T, Freedman M, Carlson M, Hill J, Frick K, Lalongo N, et al. Experience Corps: design of an intergenerational program to boost social capital and promote the health of an aging society. J Urban Health. 2004;81:94–105.
87. Rebok G, Carlson M, Glass T, McGill S, Hill J, Wasik B, et al. Short-term impact of Experience Corps participation on children and schools: results from a pilot randomized trial. J Urban Health. 2004;81:79–93.
88. Murayama H, Kondo K, Fujiwara Y. Social capital interventions to promote healthy aging. In: Kawachi I, Takao S, Subramanian S, editors. Global perspectives on social capital and health. New York: Springer; 2013. pp. 205–38.
89. Brune N, Bossert T. Building social capital in post-conflict communities: evidence from Nicaragua. Soc Sci Med. 2009;68:885–93.
90. Pronyk P, Harpham T, Busza J, Phetla G, Morrison L, Hargreaves J, et al. Can social capital be intentionally generated? A randomized trial from rural South Africa. Soc Sci Med. 2008;67:1559–70.
91. Pronyk P, Harpham T, Morrison L, Hargreaves J, Kim J, Phetla G, et al. Is social capital associated with HIV risk in rural South Africa? Soc Sci Med. 2008;66:1999–2010.
92. Fine B. They f**k u up those social capitalists. Antipode. 2002:796–9.
93. McKnight J. The careless society: community and its counterfeits. New York: Basic Books; 1995.
94. Rostila M. The social capital of welfare states and its significance for population health. In: Kawachi I, Takao S, Subramanian S, editors. Global perspectives on social capital and health. New York: Springer; 2013. pp. 277–306.
95. Muntaner C, Lynch J, Smith G. Social capital, disorganized communities, and the third way: understanding the retreat from structural inequalities in epidemiology and public health. Int J Health Serv. 2001;31(2):213–37.
96. Kawachi I, Subramanian S, Kim D. Social capital and health. New York: Springer; 2008.
97. Pollitt K. For whom the ball rolls. The Nation. April 15, 1996.
98. Gilbert K, Dean L. Social capital, social policy, and health disparities: a legacy of political advocacy in African-American communities. In: Kawachi I, Takao S, Subramanian S, editors. Global perspectives on social capital and health. New York: Springer; 2013. pp. 307–22.
99. Chwe M. Structure and strategy in collective action. Amer J Sociol. 1999;105:128–56.
100. Sobel J. Can we trust social capital? J Econ Lit. 2002;40:151.
101. Sander T, Lowney K. Social capital building toolkit, version 1.1 Cambridge, MA: Harvard University John F. Kennedy School of Government; 2005. Available from: www.ksg.harvard.edu/saguaro/pdfs/skbuildingtoolkitversion1.1.pdf.

CHAPTER 9

AFFECTIVE STATES AND HEALTH

Laura D. Kubzansky, Ashley Winning, and Ichiro Kawachi

> Let no one persuade you to cure his headache until he has first given you his soul to be cured, for this is the great error of our day in the treatment of the human body, that physicians separate the soul from the body.
>
> Socrates, from ***Charmides*** (1).

EMOTIONS AND THE SOCIAL CONTEXT

A substantial body of evidence suggests that the social environment, including conditions experienced in the family, neighborhood, and workplace, can influence health (2). Despite considerable progress in the field, how social conditions external to an individual get inside the body to influence health remains unresolved and is an active area of investigation. One proposed pathway is through emotions and the physiological, cognitive, and behavioral responses they evoke. Emotions mediate an individual's response to events in the external world, with stronger emotions evoked when events are particularly meaningful to the individual. Emotions are patterned by upstream social factors such as position in the socioeconomic hierarchy (3), and also influence individual health outcomes downstream (4, 5). As a result, emotions provide a critical window for examining the translation of conditions in the social environment into individual health status.

The social context plays an important role in determining which emotions are likely to be experienced, how they are expressed, and what their consequences will be (6). A great deal of psychological and sociological theory and related empirical evidence has suggested that emotional experience is heavily influenced by exposure to social stress (7, 8). Sources of social stress have been identified as including life events (events that lead to significant changes in an individual's life), daily hassles, role overload (occurring when demands of a role exceed an individual's capacity), or interrole conflict (defined as conflicting demands of different roles), among others.

Numerous studies have demonstrated a rise in anxiety, depression, and other forms of distress occurring in response to social stress (e.g., 9, 10–12). While proposed sources of social stress are varied, such situations are generally characterized as being threatening, unpredictable, and uncontrollable, and as overwhelming an individual's ability to cope with the demands imposed (7, 13). At the most extreme end of the spectrum of social stress is exposure to traumatic events such as abuse, maltreatment, or combat, which can result in significant emotional problems, and emotion-related disorders (e.g., post-traumatic stress disorder) (14). Kemper (6) has further suggested that emotions also arise in response to power and status differentials embedded within social situations. While this notion has sometimes been debated (15), in an experimental test of this premise, Mendelson and Kubzansky (16) manipulated social status and examined effects of subordinate versus dominant status on affective and cardiovascular responses to the experimental tasks. Compared with induced dominant status, induced subordinate status produced increased negative affect and systolic blood pressure over the course of the study. Findings provide support for the substantial observational literature suggesting that lower social status per se—above and beyond resource access—can lead to greater negative affect.

As a result, even emotions that feel highly personal and unique to the individual can be conditioned by external social factors and are therefore socially patterned.[1] Empirical support is derived from several lines of research and builds on the work discussed above linking stress and adversity to distress. Numerous studies have demonstrated that individuals with low social status more frequently encounter negative life events and chronic social stressors (18) and also may interpret ambiguous social events more negatively (19), which can lead to higher levels of social conflict (20). Additionally, previous research has found that men and women who are socially disadvantaged—of minority status or from lower socioeconomic status (SES) —generally report higher levels of distress and fewer positive emotions than do other individuals (21–23). Taken together, this research suggests that social disadvantage leads to higher exposure to chronic and acute stressors, which, in turn, negatively influence emotional experience (12).

Moreover, the larger social context influences and interacts with the family social context early in life to pattern emotion response and capacity to regulate these responses throughout the lifecourse (24, 25). Emotion regulation is learned through socialization and experience over time, and during childhood temperament, biology, and social factors interact to build regulatory skills and strategies that are then used across the lifecourse (26). Numerous studies have demonstrated that lack of warmth and high levels of conflict in the family environment are associated not only with chronic social stress but also with more emotional problems in children, as manifested by aggression, conduct disorder, anxiety, and depression, as well as other forms of distress (24). Parental ability to be sensitive to a child's emotional development, another powerful influence on whether children learn to regulate emotions effectively, is also strongly influenced by external social conditions (25, 27). More generally, chaotic and unpredictable family environments are more likely to occur among individuals with low social status; these environments are also less likely to provide the social experiences necessary for children to learn how to regulate emotions and behavior, to facilitate attachment to important individuals in their lives, or to facilitate

1 However, not all individuals in the same environment are affected by the environment in the same way. While emotions may be socially patterned, they are not determined solely by social conditions, but rather by the interaction between the individual and his or her environment (2, 17).

developing a sense of security with such individuals. Lack of appropriate social experiences in turn fosters a greater propensity to encounter social stress and to experience difficulty developing or maintaining supportive social networks. As a result individuals with these deficits are more likely to experience chronic negative affect and to be more reactive to stress across the lifecourse (19, 28). Given that they arise in response to social experience and can influence both biological function and behavior, the study of emotions may help to explain how social conditions "get under the skin" to affect health (4).

EMOTIONS AND HEALTH: A BRIEF HISTORY

Links between emotion and health have been described for over 2,000 years. For example, Rabbi Joshua ben Hananya, a first- and second-century (CE) scholar, maintained that hostility and hatred of others were among the factors that shorten the life span (Tractate Avot 215, cited in 29). In ancient times, Hippocrates considered the four bodily humors (blood, black bile, yellow bile, and phlegm) to be the basis of personality, and these elements were also believed to relate to the causes of disease (30). In 1628 William Harvey, a pioneer in cardiovascular physiology, wrote "A mental disturbance provoking pain, excessive joy, hope or anxiety extends to the heart, where it affects its temper, and rate, impairing general nutrition and vigor" (31: p. 106).

In the middle of the twentieth century, psychoanalysts suggested that psychological conflicts could trigger or contribute to disease processes, whereby somatic manifestations were considered to represent symbolic expressions of underlying repressed psychological conflicts. Particular types of conflict were linked with specific health outcomes. For example, conflicts about expressing anger were posited to lead to heart disease, or conflicts about dependency needs to ulcers (32). Empirical tests of these psychoanalytically inspired hypotheses were inconsistent at best, and research on emotion and health fell out of favor (33). However, the ideas continue to reverberate. While medical science no longer relies on the theory of humors, the basic typology has survived to some extent: We describe individuals as hopeless and depressed; as angry and hostile; as apathetic; or as optimistic (34).

PATHS LINKING EMOTION TO HEALTH AND DISEASE

With advances in research methodology and the increasing availability of prospective cohort studies in which emotion and long-term health outcomes are tracked, questions about whether and how emotions influence health outcomes have resurfaced. If emotions are posited to mediate, at least in part, the effects of social environment on health, then a critical part of the argument rests on more definitively establishing a causal role for emotions in the etiology of disease. However, important to note is that the role of emotion in disease can be considered from numerous angles. Perhaps the least controversial hypothesis is that illness causes distress and negative emotions in

various forms. Distress might in turn affect the progression or exacerbation of the illness through effects on health-related behaviors or compliance with recommended medical regimen (for more discussion of these issues see Chapter 11 this volume, and also 35, 36, 37). In contrast, the most controversial idea is that emotions actually influence the development of disease or can trigger an acute disease episode (38). The focus of this chapter will be on the role of emotions in the etiology of disease (and health), rather than as a consequence of disease or as part of the process of disease management.

Over the last several decades researchers have amassed a substantial amount of empirical evidence strongly suggesting a causal relationship between emotion and disease. The most recent evidence has suggested that effects of emotion are not uniform, such that negative emotions may increase susceptibility to disease (39) while positive emotions may protect health in a variety of ways (40, 41). Two primary pathways for such effects have been hypothesized. First, emotions may have direct physiological effects on the development of disease (or maintenance of health) via biological alterations that occur as a result either of cumulative effects of repeated emotion experiences or of an extreme and acute emotion episode. Second, emotions may also influence health by motivating (or demotivating) health-relevant behaviors such as cigarette smoking or risky sex. Chapter 13 of this volume, on behavioral economics, describes the influence of anticipatory emotions on judgments about risk (i.e., the affect heuristic). Recent research further implicates the regulation of emotion as an important determinant of the adoption of health behaviors (see later section and also, 5).

Important to note is that while there are likely distinct pathways by which emotion influences the development, triggering, exacerbation, or progression of disease, there may also be some overlap among mechanisms affecting these different stages of illness. However, effects of emotion in an already damaged biological system may be quite different from effects in an initially healthy system. Thus, careful consideration of both the affect experience (i.e., valence, intensity, duration, frequency) and the disease (type, stage, onset versus progression, severity, biological alterations) is needed when evaluating research in this area. Findings on effects of emotion on health in patient versus in healthy populations should not be considered as uniformly informing the same questions.

EMOTION THEORY: AN OVERVIEW

While theories of emotion were not developed to address the interrelations between emotions and health, consideration of the nature of emotions and their function will ultimately inform an understanding of how or why they influence health. Emotion theorists have suggested that emotions may be conceptualized as having cognitive, neurobiological, and behavioral components (42, 43). Specific emotions are thought to be biologically based, arising as a product of the interaction between the person and the environment, and mediating between continually changing situations and the individual's behavior (44, 45). The state of subjective feeling serves as a compelling signal that the person is faced with a particular type of challenge, and motivates the person to respond to this fact (43). For example, fear motivates a person to escape danger, sadness motivates a person to disengage from loss, and so on (17). Thus, even negative emotions are considered functionally

appropriate processes; however, they may have dysfunctional consequences when the organism is taxed beyond the limits of its capability (46).

Each emotion depends on an individual's appraisal of events in terms of their importance and the demands they place on the individual, as well as in terms of the individual's options and prospects for coping (17). More specifically, individuals appraise events in terms of whether they are potentially harmful (e.g., associated with threat or loss) or beneficial (e.g., associated with actual or potential gain) (17, 47). Emotions also serve to communicate a person's emotional state and likely behaviors to others in the social environment (43). And, emotions are associated with urges to act in particular ways called action tendencies, which enable the individual to cope with environmental demands (43, 48). That said, while particular urges to act may be associated with specific emotions, people do not invariably act out these urges when experiencing particular emotions.

Most emotions may be seen either as transitory *states* brought on by specific situations, or as *traits*, that is, stable and general dispositions to experience particular emotions (49).[2] Individuals high in trait anger, for example, experience the transitory state of anger more frequently and intensely than individuals low in trait anger. Thus, certain personality types are hypothesized to be vulnerable to disease in part because these individuals are predisposed to experience particular emotions (53). For example, hostility is considered a personality trait that predisposes individuals to experience more episodes of anger, suspicion, and cynicism than other individuals. Because hostile individuals may create hostile environments by engaging in cynical, mistrusting, and aggressive behavior, they create more opportunities to experience anger (54).

A separate line of theoretical work has specifically sought to understand positive emotion. The broaden and build model (48), a dominant theoretical perspective, suggests that positive emotions actually produce optimal functioning. This model identifies four positive emotion families including *joy, interest, contentment*, and *love*, and proposes that positive emotions lead to a broader range of thought and action tendencies. Broadening these tendencies over time serves to build personal resources. For example, joy creates the urge to play and be creative while interest creates the urge to explore. Play may help to build social resources by strengthening social bonds, while exploration increases knowledge and intellectual complexity. In this way, positive emotions facilitate successful adaptation to ongoing demands.[3]

More generally, negative and positive emotions and the interplay between them are products of an *emotion regulation* process, which involves monitoring and managing one's emotional experience and response (56). Emotion regulation is learned through socialization and experience over time, and therefore is likely heavily influenced by the social environment. As a result, investigators have increasingly focused on the role of emotion regulation as an aspect of "self-regulation"—together with the ability to focus attention, resist impulses, and delay gratification—that is tied to the functioning of the prefrontal cortex, or what behavioral economists refer to as the function of "System 2" in monitoring the impulses of "System 1" (the dopamine reward centers of the brain) (see Chapter 13, on behavioral economics). Early theories speculated that aspects of

2 Emotions are considered psychological entities separate from moods or attitudes. Emotions are generally considered to have an object so that they are "about" something, whereas moods have been defined as being more diffuse, lower in intensity and of longer duration than emotions (50). Emotions are one component of attitudes, which have been defined as learned predispositions to respond in a consistent manner with respect to a given object (51, 52).

3 It is perhaps worth noting that much of the research on positive states of mind and health has blurred distinctions between positive emotion and other types of positive psychological factors such as optimism (55).

emotion regulation might be linked to health because the effort involved in suppressing or inhibiting emotions was posited to be biologically costly, thereby leading to increased susceptibility to illness over time (57, 58). More recent formulations have moved away from defining regulation as expression versus suppression and suggest that emotion regulation is a dynamic process involving both up- and down-regulation of positive and negative emotions (59). Some strategies may be considered antecedent-focused, employed *before* an emotion occurs; for example, in the context of a stressor individuals could reappraise the situation, changing their cognitive appraisals in a way that might prevent or reduce the intensity of subsequent negative emotions. In contrast, response-focused strategies are employed *after* an emotion has occurred and involve modifying the behavioral manifestations of the emotion (e.g., suppression); such efforts may be taxing and may not mitigate the negative emotional experience (60). The appropriateness of any given strategy is context-dependent, although some research has suggested that greater reliance on antecedent-oriented regulation may be more adaptive, particularly in the context of health (61).

APPROACHES FOR INVESTIGATING EMOTION AND HEALTH

MODELS OF EMOTION AND HEALTH: MECHANISMS AND UPSTREAM DETERMINANTS

Emotions are hypothesized to influence health directly because they evoke physiological processes (e.g., activation of the hypothalamic-pituitary-adrenal [HPA] axis and the sympathetic nervous system [SNS]), and indirectly because they influence health-relevant behaviors via motivation and decision-making (see Figure 9.1). Pathways by which negative emotions may directly alter biological processes have been identified and evaluated in both animal models and human populations. For example, elevations in serum norepinephrine levels associated with negative emotions may increase blood lipids, free fatty acids, blood pressure, and heart rate, and lead to constriction of peripheral blood vessels. Negative emotions such as anxiety and depression may also lead to altered autonomic regulation of the heart (e.g., 62). Other direct biological effects of emotion on health may be through altered immune functioning. Immune cells have receptors for cortisol, epinephrine, and norepinephrine. Thus, activation of the HPA axis and the SNS, which results in elevated serum levels of cortisol and catecholamines (63), can also lead to dysregulation of immune function (64, 65). Over time, recurring activation of these systems may set disease-related physiological processes in motion. In fact, there is a growing body of empirical evidence linking negative emotions like depression and anxiety with chronically elevated levels of inflammation (e.g., 66, 67). Direct biological effects of positive emotions are less well studied, but emerging evidence points to possible salutary effects on lipids (68), inflammation (69), and vagal tone (70).

Epidemiologic studies have also demonstrated that behavioral risk factors (smoking, alcohol consumption, physical activity, body mass index) mediate relationships between trait negative emotions like anger or depression and cardiovascular morbidity and mortality (see Chapter 10; 71). State emotions may also indirectly affect health through associations with behavioral and other

risk factors. Positive emotions such as pleasure and joy are routinely targeted in advertisements for harmful products (such as cigarettes) with the aim of lowering consumers' judgments about the risks associated with their use. Conversely, negative emotions such as anxiety and depression may activate dopamine reward centers of the brain among smokers, making it more likely they will reach for another cigarette (e.g., 72). Emotions also influence social processes including the quality and quantity of an individual's social relationships as well as cognitive processes including health-related decision-making, which in turn influence health (see Chapters 7, 13; and 5, 73).

Much of the literature on emotion and health relates to the mechanisms connecting emotion to physiology and pathology, but an important task for social epidemiologists is to put emotions in a social context. In fact, while investigators commonly posit stress and emotion as likely mediators explaining the effects of social disadvantage on health, studies directly testing this hypothesis remain surprisingly limited, perhaps due partly to challenges in the data and methodological requirements of such studies (4). The limited empirical work notwithstanding, a more nuanced consideration of how the social environment may pattern emotions has emerged over the last decade. Based in part on the observation that socially disadvantaged individuals tend to be not only exposed to more potential stressors but also more reactive to them, the reserve capacity model has been proposed (4). The model posits that individuals of low SES achieve and maintain a smaller reserve of resources that promote effective management of stressful circumstances and reduce likelihood of experiencing significant distress. These resources may include social resources (e.g., social support, neighborhood social capital) and intrapersonal resources (e.g., sense of control, self-esteem). Some researchers also consider "willpower" or, more generally, the capacity to self-regulate as a form of reserve capacity (74). Self-regulation—for example, the ability to resist temptation and to delay gratification—tends to become drained under duress. Self-regulation has also been linked with physiologic reserve. For example, alcoholics have lower vagal tone (as assessed by lower heart rate variability) compared with normal controls (75). Among alcohol-dependent patients treated at a long-term residential facility, post-treatment high-frequency heart rate variability (HRV, indicating vagal tone) significantly predicted the occurrence and timing of drinking relapse by 6-month follow-up, independent of alcohol dependence severity (76). Higher HRV has thus been put forward as a biomarker of regulatory capacity reserve (77). Moreover, HRV appears to be strengthened just as capacity to self-regulate can be strengthened, like a muscle, by controlled breathing and meditation (78). The ability to develop such reserves among low SES individuals may be limited, however, because individuals in disadvantaged circumstances are exposed to more situations that require use of their resources, thereby depleting their reserves, resulting in what Sendhil Mullainathan and Eldar Shafir have labeled the "bandwidth tax" (79).

Figure 9.1 presents a heuristic model designed to illustrate the links between the social environment, emotion, and health. For the purposes of parsimony we have presented a unidirectional model; however, the exclusion of alternative paths is not intended to reflect hypotheses about their existence or to imply these relations are static.

STUDY DESIGN AND METHODOLOGICAL ISSUES

Research on the direct health consequences of emotion states has typically focused on the immediate physiological responses associated with emotion experiences. In contrast, research on emotion

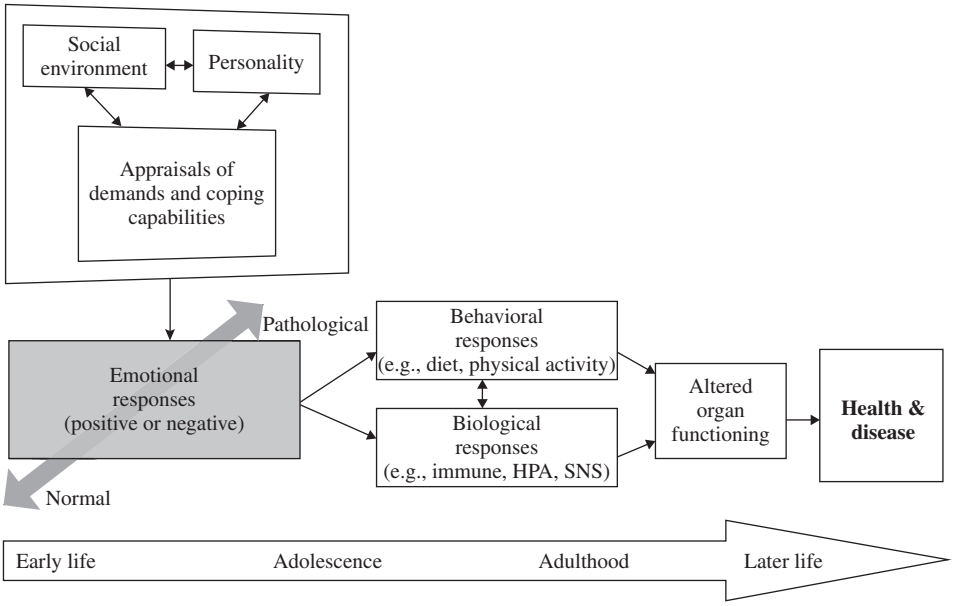


FIGURE 9.1: A model of the stress-emotion-health process.

traits tends to examine the long-term health effects of recurring emotion experiences. Short-term effects of acute emotion states are generally examined in the laboratory to identify their direct effects on physiological parameters hypothesized to be related to disease processes. For example, research on cardiovascular reactivity has measured an individual's propensity to react to acute stress with increased heart rate and blood pressure (e.g., 80). It has generally been assumed that measurable short-term physiological effects of emotions are related to their long-term health consequences (81), and a recent meta-analysis provides support for this assumption. This meta-analysis reviewed research looking at cardiovascular responses to laboratory-induced distress in relation to future cardiovascular risk status; overall, greater reactivity to and poorer recovery from stress were prospectively associated with worse cardiovascular outcomes, including hypertension and higher atherosclerotic burden (82). Additional studies published since the meta-analysis are consistent with this conclusion (e.g., 83). Other laboratory and experimental research has found adverse acute effects of emotion on other biological parameters related to numerous health outcomes, such as natural killer cell cytotoxicity, circulating inflammatory markers (e.g., C-reactive protein), and cortisol (e.g., 84, 85).

Examining effects of acute emotions on the triggering of disease has been done primarily in the context of coronary heart disease (CHD; e.g., acute myocardial infarction and takotsubo cardiomyopathy). Challenges revolve around assessing acute emotion states in relation to disease onset without inducing confounding due to retrospective reporting bias (for detailed discussion of methodologic issues see 86). The case-crossover design, in which exposure to a transient risk factor is contrasted with exposure during a "control" period within the same individual, mitigates this concern somewhat. Studies with a case-crossover design provide some of the strongest evidence that acute episodes of anger, anxiety, and depression may trigger acute coronary events (87, 88).

By contrast, more traditional case-control (cross-sectional) and longitudinal study designs have been used to examine the relationship between stable trait emotions and incidence of disease. Although emotions are usually measured at a single point in time, the measurement approaches used in these studies are designed to identify the chronic nature of the emotion experience.[4] Randomized controlled trials have been conducted to assess more rigorously whether emotion is causally associated with increased risk of disease; however, due to logistic and feasibility issues, trials are conducted in patient as opposed to healthy populations (e.g., trials on depression and CHD have studied the relationship among individuals who have had an acute coronary event). Disease-related processes may differ depending on whether individuals are healthy or not at the outset. Given the lack of feasibility of true experiments in which people can be assigned to experience one emotion or another prior to disease development, prospective cohort studies present the strongest evidence for the hypothesis that emotions influence health. The most convincing design is to measure emotions among initially disease-free individuals, thereby preserving the temporal order of the linkage between emotion and disease onset. However, these designs remain susceptible to concerns about hidden sources of reverse causality (i.e., predisease states influence emotions) and the possibility that some unmeasured variable (e.g., a genetic risk factor) may underlie the apparent emotion-health relationship.

DEFINITION AND ASSESSMENT

DIFFERENTIATING BETWEEN STRESS AND EMOTION

Stress has often been proposed as a way social conditions may get inside the body. External circumstances or events, characterized as environmental demands or stressors, are hypothesized to cause psychological and/or physical stress, which in turn may cause physiological changes related to disease processes. This conceptualization has its origins in physics, which defined stress as the force exerted on a material structure, leading to structural strain, damage, and potentially collapse if the load is more than the structure can bear (5). In the context of health, most stress theories assume that stress is harmful because repeated experience leads to the accumulation of damaging physiological changes. The physiological basis of stress was first proposed by Hans Selye, who suggested that physical and psychosocial stressors both elicit the same pattern of physiological response (89).

Early tests of the stress-health hypothesis examined the health effects of the accumulation of stressors of high magnitude characterized by life events (e.g., moving house, birth of a baby), and of lower magnitude, characterized as daily hassles (chronic but lower intensity stressors such as concern about paying bills) (90, 91). While findings supported some of the expected associations, they were not as reliable or robust as expected. One problem with the original theory is that no distinction is made between stressors in terms of the physiological strain they impose. For example,

4 The ongoing interaction between emotion and health may be particularly difficult to capture in research investigations that measure emotion at a single point in time. As a result, reports of the relationship between emotion and health may be underestimates.

in this formulation, prolonged exposure to a loud noise is considered functionally similar to experiencing the death of one's spouse. The theory also focuses primarily on the physiological set of reactions elicited by noxious stimuli (external stressors), and does not address the psychological appraisal of the stimulus. Moreover, the theory cannot account for individual differences in reactions to stressors. In fact, the physics origins of the conceptualization of stress suggests that failing to take account of both load on and capacity of the system is problematic. Thus, later formulations of stress theory identified two components as critical in determining the potential health effects of a stressor. The first is the magnitude of the stressor (conceptually similar to the weight of the load) based on the assumption that stressors can be reliably characterized as ranging from "small" to "large"; the notion here is that "large" stressors create more strain than "small" stressors. The second component relates to the capacity of the individual (conceptually similar to the hardiness of the structure experiencing the load). The assumption here is that individuals appraise the "same" stressor differently, with some individuals unaffected by stressors that lead others to collapse.

Following this line of thinking, investigators put forth a more psychologically oriented theory of stress to explain *when* an individual will experience stress, and link psychological processes to the physiological processes described by Selye (7). In this formulation, stress is experienced when individuals perceive (appraise) that external demands exceed their ability to cope. The interpretation of an event as stressful triggers a series of physiological changes. While this is a conceptual improvement, empirically it has proven challenging to operationalize. It is difficult to predict when demands will exceed an individual's capacity to cope or which individuals will have sufficient coping capacity and under what circumstances. Absent measuring this capacity, it may appear that some individuals undergo many stressful events with few health consequences while others have seemingly trivial problems, but experience poor health outcomes.

Defining stress has also proven difficult, as the term is used broadly. For example, it is unclear exactly what it means when an individual reports experiencing stress. Early stress theorists including Selye (92) proposed different kinds of stress, such as good stress termed "eustress" (e.g., planning a wedding) versus bad stress (e.g., death of a spouse). Investigators have tried to define stress objectively by adding up the number of potentially stressful events or daily hassles people experience. However, what one individual considers stressful, another may not. Thus, simply summing numbers of "stressful" events reported by an individual may not reveal much about an individual's life experience. To know if someone is "stressed" we need to know the individual's interpretation of the potentially stressful event and its meaning for his or her life. Because emotions can provide some important clues about whether the stressor was appraised by any given individual as bothersome, Lazarus (93) and others have suggested focusing more strongly on the study of emotion.

While there is some overlap between stress and emotion and their relationships to health, there are also important distinctions between them. For discussion in this chapter, environmental events are considered stressors while emotions are considered responses to stressors (73). A negative emotion response typically occurs if demands are perceived to exceed one's ability to cope. Theories of stress and emotion are relatively silent on positive emotional responses and when they might occur (17, 94) and less clear on the role of emotional traits. Also important to note is that potentially stressful events can be associated with a variety of different emotions. For example, losing one's job may provoke anger in some individuals and depression in others. As emotions can be considered products of stress as well as mediators of its effects (95), they may provide a more nuanced way of understanding an individual's interaction with the environment.

PATHOLOGICAL VERSUS NORMAL EXPERIENCES OF EMOTION

The experience of most emotions occurs along a continuum. There is a range within which emotion levels are considered to be normal; but when they occur in inappropriate contexts and/or at high intensities, they may be identified as pathological (46). Anxiety and depression are commonly experienced emotions that can also underlie clinical disorders. For example, an anxiety disorder is considered to be present when the experience of anxiety is *(1)* recurrent and persistent; *(2)* of an intensity far above what is considered reasonable, given the objective danger or threat; *(3)* paralyzing so that individuals feel helpless and unable to cope; and *(4)* the cause of impaired psychological or physiological functioning (96, 97). Psychological research has suggested that pathological anxiety (clinically diagnosed conditions like panic disorder or generalized anxiety disorder) and normal anxiety reactions are essentially similar in their cognitive, neurobiological, and behavioral components (46). Thus, anxiety can refer to both the normal and pathological spectrum of symptomatology (97, 98). While psychiatric epidemiologic research has largely focused on mental health disorders as primary outcomes, the epidemiologic research on emotions and health to date has considered both subclinical (in the range of normal) and more pathological manifestations of emotion (99). Health effects are evident across the spectrum, and emotions appear to have a dose-response association with risk of disease onset. Thus, it appears that effects of emotion on health are relevant for a broad group of individuals in the population, not simply for those with clinically relevant levels of emotional problems.

MEASUREMENT ISSUES

Most epidemiological studies on emotion and health have relied on self-reports of emotion, using a specific emotions approach. This approach theorizes that there are many different types of emotions, each with different characteristics and specific response patterns (100, 101). Scales with labels like "happiness," "sadness," and "anxiety" have been derived using this approach. Measures in this tradition generally include adjective checklists or lists of statements that respondents endorse in terms of the extent to which each statement applies to them.[5] A broad array of reliable and well-validated scales of emotions exist for measuring anger, anxiety, and depression (see for example 103, 104), but fewer for positive emotions. Of note is that epidemiologic studies often consider depression and anxiety as representing single emotions; in contrast, psychologists have argued that while these states are commonly characterized by dysregulated emotion, they also reflect complex constellations of chronic elevations of maladaptive cognitions and behaviors (105). However, given the primacy of the emotion component and the insight that an emotion-oriented framework may bring to understanding their role in health, for the purposes of this chapter we refer to each of these states simply as an "emotion."

5 In contrast, a dimensional emotion approach builds on the notion that there are a small number of dimensions that describe all emotions (i.e., pleasantness, activation), and specific emotions are derived from combinations of these basic dimensions (102). Research on emotion and health, however, benefits from the specific emotion approach, because dimensional approaches miss much of the richness of affective life and do not convey differences in how different emotions are experienced in physiological and behavioral domains (17).

Self-report assessments have a number of problems. Study participants must be willing to disclose what may feel like private information. Some participants may want to present themselves in the best light possible, and therefore fail to respond accurately to the questions (a phenomenon called social desirability). Thus, self-report data may not distinguish between genuine mental well-being (i.e., low anxiety) and the façade of health created by psychological defenses (57). Moreover, individuals may lack insight into themselves and fail to give accurate reports of the emotions they experience. A further issue with self-report is what to include in the domain of emotion. Many scales ask about somatic symptoms (e.g., racing heart), since emotions in general often co-occur with physical symptoms (106). As a result, using emotion scales that include symptom assessments to predict certain types of health outcomes (e.g., general symptomatology) may be misleading. Scales should be carefully screened on whether their content is appropriate for the study being undertaken. Measures other than self-report are possible, each with its own benefits and pitfalls. These include peer report, such as asking spouses to provide emotion ratings for study subjects, and observer ratings, in which trained interviewers observe and rate subjects on their emotions.

Classic epidemiologic methods seek to characterize individuals according to whether or not they are exposed to the risk factor of interest. Whatever type of emotion measure is used, it may be difficult to define someone as "not exposed" to emotions like anxiety or anger or happiness, since almost everybody generally experiences some level of each emotion. Even the definitions of emotionally based psychiatric disorders are not constant. Current diagnosis of mental disorders is based on clinical observation and phenomenological symptom reports by patients. However, there is increasing recognition that the existing classifications may not capture underlying biological specificity and that new classifications may ultimately be necessary (107). Many epidemiologic studies to date have been opportunistic, making use of whatever measure may be available, recognizing that the items commonly available (often a single item or a set of items that have not been evaluated for their psychometric quality) may not be ideal. In going forward with research designed at the outset to consider emotion, it will be important to consider carefully the instruments used to measure each emotion (for more detailed discussion of these issues, see 102), as different measures may be appropriate depending on the hypothesis being investigated.

EMOTION INTERRELATIONSHIPS

Considerable overlap occurs in the components of various emotions. For example, appraisals of threat may be common to both anxiety and anger, while both anxiety and elation may be accompanied by cardiovascular arousal. How do these similarities affect the ability of researchers to detect specific health effects of a given emotion? Prior work has suggested there may be specific physiological patterns associated with each emotion (108), but the reliability of these differences has been debated (109, 110). In part, this may be because investigations to date have looked at only a limited array of physiological parameters. In fact, even among emotions that seem to call forth similar physiological responses (e.g., heightened autonomic arousal) such as anxiety and elation, there are some distinctions. For example, in the presence of a stressor, negative emotions are associated with appraisals of threat while positive emotions are associated with appraisals of challenge

(111, 112). Laboratory studies have demonstrated that individuals who feel threatened by a stressful task exhibit less cardiac reactivity (measured by heart rate, cardiac contractility, cardiac output) and higher vascular resistance while those who feel challenged exhibit greater cardiac reactivity and decreased vascular resistance (112). With increasing technical sophistication in measuring biology using the rapidly expanding capabilities in genomics, cell biology, and measuring circuitry in the brain, there is a renewed sense that neurobiological specificity exists and can be identified (107).

Whether or not specific emotions have a distinctive neurobiological profile, the behaviors they motivate are often quite different and separable, and thus effects on health may differ regardless of the underlying neurobiology. For example, anxiety is associated with active efforts to cope with difficult situations and heightened vigilance, whereas depression is more often characterized by behavioral retardation and withdrawal (42, 113). Similarly, anger is often associated with an impulse to approach others and be aggressive while anxiety is more often associated with vigilance or the desire to escape. These behavioral patterns may have important implications for health, as they may differentially influence the likelihood that individuals subsequently engage in health-promoting or impairing behaviors. Indeed, appraisal tendency theory (114) suggests that emotions with the same valence (e.g., fear versus anger) can lead to differential judgments about risk as well as opposing action tendencies. Thus the elicitation of fear (e.g., via graphic warning labels on cigarettes) leads to a heightened assessment of the risk associated with the behavior (in this instance, smoking), whereas anger appears to lead to a more optimistic appraisal of the future. The field of behavioral economics is only beginning to explore the specificity of emotions in influencing consumer judgment and choice and hence the adoption of various health behaviors—see Chapter 13, this volume for more discussion of this topic.

Behavior is commonly referenced as an important pathway by which emotions may influence health; however emotions are rarely considered as modifiable upstream determinants that may serve as important points for intervening to improve behavior. Despite progress, strategies for nudging behavioral choices through the differential elicitation of emotions remain limited and there is a pressing need to refine existing approaches or develop new ones (115). New traction may be gained by bringing in a stronger and more nuanced emphasis on emotions that may motivate or demotivate health behaviors, moving beyond the knee-jerk use of fear appeals in public health messaging campaigns to motivate healthy behaviors.

Despite the potential for specific emotions to influence distinct behaviors, specificity of emotion/disease associations may yet be difficult to establish because emotions rarely occur in isolation (104, 105). For example, anxiety and depression often occur together, but studies that examine the association of anxiety with health outcomes often fail to account for their overlap (116, 117). More recent studies (largely in the realm of CHD) have begun to try to tease apart emotion-specific effects with some success. For example, a study of Vietnam veterans examined the effects of depression and anxiety (characterized by psychiatric diagnosis) and their co-occurrence in relation to cardiovascular disease mortality (118). While each emotion was separately associated with increased risk, high levels of both emotions together conferred the greatest risk for earlier death over the follow-up period. Given the often high level of overlap between some of the emotions of interest, such work will benefit from innovative methodological approaches that go beyond standard statistical methods for adjusting for confounding. However, there is currently enough evidence of separable effects to suggest it is

prudent to continue to consider specific emotions separately, while acknowledging and actively investigating their important shared components (119).

THE EPIDEMIOLOGICAL EVIDENCE ON EMOTION AND HEALTH

Much of the research examining the role of emotion in maintaining health or in the etiology of disease has been carried out in the context of cardiovascular disease (CVD) with a particular focus on CHD (i.e., myocardial infarction [MI], sudden death, angina). This research emphasis is due to a number of factors. Cardiovascular disease is the leading cause of death globally, with more people dying annually from CVD than from any other cause (120, 121), and as a result it is a frequently measured outcome in epidemiologic research. Coronary heart disease has been a particular focus because onset, triggering, and exacerbation can each be more clearly identified for CHD than for other diseases and also other risk factors are well known and can be carefully considered.

While current evidence suggests that traditional risk factors (e.g., smoking, hypertension, hypercholesterolemia, obesity, physical activity, diet, diabetes) explain a substantial fraction of the occurrence of CHD, these factors primarily explain the final common causal pathway for the disease (122). Because major population trends in the occurrence of the disease are due in large part to environmental factors, calls for research to explore the broader social and economic determinants of these risk factors have been increasingly common (122). In addition, there has been an emerging emphasis on primordial prevention, which has the goal of preventing whole societies from developing an epidemic of a risk factor, or correspondingly of preventing the development of risk factors within an individual (123, 124). Together these trends have led to renewed interest in how emotions (and other psychosocial factors) influence the development of CHD and related cardiometabolic conditions.

Evidence for associations between emotions and the development of other health outcomes tends to be more sparse. As a result, in the sections to follow, the emotion-CHD relationship is emphasized, recognizing that research with CHD can serve as a strong model (albeit likely incomplete) for understanding the role of emotions in health more broadly. In describing research on emotions and other health outcomes we will highlight what has been done and give some consideration to potential barriers for this research. We will subsequently discuss exciting directions for future research.

CORONARY HEART DISEASE

In the 1950s, two cardiologists, Friedman and Rosenman, proposed a new risk factor for CHD called the Type A behavioral (TAB) pattern (125). The TAB pattern was characterized as an action-emotion complex that requires an environmental challenge to serve as the trigger for expression. The overt manifestations of the behavior include a free-floating but well-rationalized hostility, hyperaggressiveness, and a sense of time urgency. Several large-scale epidemiological studies conducted during the 1960s and 1970s appeared to corroborate the Type A hypothesis,

culminating in a National Institutes of Health panel that concluded in 1981 that TAB was an independent risk factor for CHD (126). However, enthusiasm for the TAB concept started to wane in the mid-1980s following the publication of a series of cohort studies that failed to find a relationship with CHD (for excellent review see 127, 128). The conflicting evidence on TAB and CHD may be partly due to the fact that the self-rated TAB questionnaires used in negative studies did not inquire about the full range of behaviors associated with the action-emotion complex (129). The videotaped structured interview is regarded as the most sensitive approach to diagnosing the TAB pattern, but it has obvious limitations in the context of large-scale longitudinal studies. Although a meta-analysis of 18 controlled trials concluded that psychological treatment for TAB resulted in a 50% reduction in recurrent coronary events (130), the focus of research on psychological predictors of CHD has gradually shifted away from TAB toward examining the relationships between specific negative emotions and CHD (131).

Given the similarities between certain emotional states, investigators hypothesized that a number of negative emotions may be risk factors for CHD (132), and research has generally corroborated this hypothesis. Links between risk of incident CHD and specific emotions like anxiety, anger, and depression have emerged, with the volume of research on the topic increasing dramatically in the last decade. As a result, numerous reviews and meta-analyses have now been published on depression (e.g., 133, 134), with markedly consistent findings, with meta-analyses also available for anxiety (e.g., 135) and anger (e.g., 136). Research on positive emotion has also increased exponentially (40), though research on emotion regulation per se remains somewhat limited (but for detailed review of the findings, see 61).

Given the volume of work now available on the topic, in the following sections we briefly describe the research on these emotions, focusing on landmark studies, and key findings (137). We prioritize prospective studies designed to look at incidence of CHD, considering only "hard" disease outcomes (e.g., nonfatal MI, sudden death) in population-based samples with individuals who are either disease-free at the start of the study or whose initial health status is controlled in statistical analyses. We also prioritize studies that control for a broad range of coronary risk factors, including health behaviors such as smoking. However, we note that if such behaviors are believed to be on the pathway between emotion and CHD, then risk estimates derived from these studies are almost always going to be underestimates. We also describe current findings in the research on emotion regulation and on acute emotions in relation to incident CHD.

ANGER

Chronic anger and hostility have long been implicated in the etiology of CHD (29). Anger and hostility are strongly associated with one another and have been implicated as "toxic" components in the relationship between Type A behavior pattern and CHD (128, 138). Hostility is a long-standing attitudinal disposition, as opposed to anger, which is considered an emotion and a component of hostility (139). Anger is defined as an unpleasant emotion arising in response to events that are perceived as unjust, and accompanied by physiological arousal and the activation of action tendencies or impulses toward aggression. Research in this area has focused somewhat more on hostility rather than on anger per se (for comprehensive review of this work, see 136).

Several meta-analyses have provided evidence that chronic anger and hostility are independent risk factors for the development of CHD (136, 140). In an early study, Kawachi and colleagues (141) examined the association between anger and CHD in a 7-year follow-up of 1,305 men in the Normative Aging Study. Compared with men reporting the lowest levels of anger, the relative risk (RR) for men with the highest levels of anger was 2.66 (95% confidence interval (CI): 1.26–5.61) for incident coronary events (including nonfatal MI, fatal CHD, and angina pectoris). A dose-response relation was found between level of anger and overall CHD risk, even after relative risks were adjusted for other major cardiovascular risk factors. A 6-year follow-up study of 12,990 middle-aged men and women (Black and White) also found that individuals with a strong tendency toward quick, minimally provoked or unprovoked anger had significantly increased risk of acute MI and fatal CHD, with a multivariate adjusted hazard ratio (HR) = 2.28 (95% CI: 1.29–4.02) (142). However, not all studies find an effect of anger, and in fact, the overall effect size in the most recent meta-analysis was modest (HR = 1.19, 95% CI: 1.05–1.35), with effects no longer significant when considered only among studies that adjusted for key covariates (136). As a result, the authors suggested effects of anger on CVD are primarily mediated by behavioral pathways.

ANXIETY

Because the word "anxiety" represents both a lay construct and a scientific term, confusion has frequently arisen about the precise meaning of the term (143). Anxiety has been defined as a future-oriented negative affective state resulting from perceptions of threat, characterized by a perceived inability to predict, control, or obtain desired results in upcoming situations (42). The most recent meta-analysis of 20 prospective studies concluded that anxiety is an independent risk factor for CHD with a pooled HR for incident CHD of 1.26 (95% CI: 1.15–1.38) and a somewhat weaker HR reported when considering only five studies that isolated effects on nonfatal MI (135). Five of the studies also adjusted for depression, and in four of them an independent effect of anxiety was maintained. The Northwick Park Heart Study, one of the earliest studies to consider this association using a prospective design, followed 1,457 initially healthy men for a period of 10 years and reported a striking association between self-reported symptoms of phobic anxiety and fatal CHD (144). Compared with men with the lowest level of anxiety, those with the highest levels had a relative risk of fatal CHD of 3.77 (95% CI: 1.64–8.64) and the association persisted after controlling for a number of cardiovascular risk factors (144). Findings have been replicated in women. For example, in a substudy within the Women's Health Initiative Observational Study, 3,369 community-dwelling, healthy postmenopausal women reported on their experience with panic symptoms over the prior 6 months and were then followed for an average of 5 years (145). Women reporting at least one full-blown panic attack were at approximately 3 times the risk of developing CHD or a stroke (relative risk [RR] = 3.08, 95% CI: 1.6–5.94) with increased risk still evident (albeit of lower magnitude) among women reporting less severe panic attacks. Findings were maintained after adjusting for standard coronary risk factors as well as for depression.

Recent research has included investigations of post-traumatic stress disorder (PTSD), formally classified in psychiatry as an anxiety disorder. Studies have consistently identified increased

risk of developing CHD among individuals with PTSD or even among those with subclinical levels of symptomatology (146). For example, in the largest study to date, 39,324 World Trade Center Health Registry participants were followed prospectively for an average of 2.9 years (147). Men and women who reported PTSD at study enrollment had elevated risk of developing CHD over the follow-up (HR = 1.68, 95% CI: 1.33–2.12 women; HR = 1.62, 95% CI: 1.34–1.96 men). Findings were maintained after adjusting for standard coronary risk factors as well as relevant 9/11 exposures such as injury or dust cloud exposure. Thus, as numerous reviews have recently concluded, anxiety in a variety of forms appears to be consistently associated with elevated risk of developing CHD (119, 148).

DEPRESSION

For an extensive discussion of depression and its relationship to cardiovascular disease, readers are referred to (35). Here we briefly highlight only the most relevant research on the role of depression in CVD incidence. The majority of the research on negative emotions and CHD has focused on depression as a potential risk factor. In an early prospective study, Anda and colleagues (149) examined the relationship between depressed affect and ischemic heart disease (IHD) incidence in 2,832 healthy adults in the National Health Examination Follow-Up Study. Depressed affect was associated with a significantly increased risk of fatal (RR = 1.5; 95% CI: 1.0–2.3) and nonfatal (RR = 1.6; 95% CI: 1.1–2.4) IHD. In 2007, a meta-analysis including 28 published studies demonstrated a positive association with incident CHD, with an RR of 2.54 (95% CI: 2.07–3.10) for individuals with clinically relevant depression levels and an RR of 1.39 (95% CI: 1.26–1.54) for individuals with depressed mood (150). Prior meta-analyses, as well as additional prospective studies, have been conducted, and most report significant associations of similar magnitude (e.g., 133, 151, 152). A recent meta-analysis of the relation between depression and incident stroke also reported associations of similar magnitude, reporting a pooled adjusted HR of 1.45 (95% CI, 1.29–1.63) (153). This estimate is consistent with other studies published since then (e.g., 154), although recent work has suggested the association is primarily evident among younger individuals (aged < 75 years; 155). For both CHD and stroke, few studies have found evidence of a threshold effect; rather studies more commonly find a dose-response relationship whereby risk is increased as depressive symptoms increase.

Other more severe instantiations of depression have also been linked with increased risk of CVD. Some studies have found strikingly stronger relative risks with use of antidepressant medication (e.g., 156) leading to speculation that antidepressants themselves may in part be driving the effects; however, other studies have more directly tested whether risks appear to be due to or exacerbated by antidepressants and have reported weak or null evidence in support of this hypothesis (157, 158). In the study with the most detailed information on depression treatment, findings indicated reduced risk associated with a range of antidepressants including selective serotonin reuptake inhibitors (SSRIs) and tricyclics (158). At present, research suggests that use of antidepressants does not confer additional risk per se but rather serves as an indicator of more severe depression. However, it is important to recognize that none of these studies were randomized controlled treatment trials designed to look at this question specifically.

POSITIVE EMOTION

Recent epidemiologic work has increasingly suggested that positive emotions might protect against CHD (for comprehensive review, see 40). Using rigorous methods, including healthy populations at the outset of each study, psychometrically valid evaluation of positive affect, and adjustment for an array of potential confounders including negative affect, studies have consistently found reduced risk of incident CHD in relation to positive affect (40). For example, several population-based studies have considered emotional vitality, a composite measure that captures a sense of interest, enthusiasm, and capacity to regulate emotion, in relation to incident CHD and found 20–30% reduction in risk after adjusting for known potential confounders as well as psychological ill-being (159, 160). In another study, positive emotions were assessed based on structured interviews and evaluations of positivity displayed on participants' faces during these interviews (161). In this sample of 2,000 men and women, those who displayed more positive affect were at 22% reduced risk of developing heart disease over a 10-year period, after controlling for major coronary risk factors and measures of negative affect. Investigators have argued for continued efforts to understand the role of positive functioning in health, suggesting that such efforts will provide enhanced understanding of how mental and physical health processes interact, and give greater insight into how to build resilience (40).

EMOTION REGULATION

With the recognition that negative affective states are characterized in part by dysregulated emotion, an early line of research hypothesized that the certain strategies commonly used to regulate emotion may be cardiotoxic (58, 162). Inhibiting expression of emotion may also impair symptom recognition, delay help-seeking behavior, and compromise communication about problems and concerns (163). While it can be difficult to assess whether individuals are suppressing or inhibiting emotions, some empirical work has provided suggestive evidence that aspects of regulation matter for health. For example, in the Framingham study, the single item "inability to discuss angry feelings" was associated with subsequent CHD risk (164).

In recent work, researchers have begun to consider the regulation of emotions more broadly as a higher order feature of emotional functioning that might help to explain diverse findings linking both positive and negative emotions with cardiovascular outcomes (for more detailed review, see 61). For example, a prospective study of 1,122 older male participants considered the relation between self-regulation and the development of CHD (165). Self-regulation was assessed according to the men's ability to manage impulses, feelings, and behaviors, with emotion regulation identified as a central feature. Compared with men who had the lowest levels of self-regulation, those with the highest levels had 62% reduced risk of experiencing a nonfatal MI or CHD death over 13 years of follow-up. Findings were maintained after adjusting for known coronary risk factors as well as main effects of positive and negative affect. A study of 7,933 Finnish adults that measured levels of anger expression and control found that participants reporting the lowest levels of anger control had 35% higher risk of experiencing a fatal or nonfatal cardiovascular event in the subsequent 10–15 years as compared with those with the highest levels of anger control (166). Findings

were maintained even after taking account of standard coronary risk factors as well as depressive symptoms.

Another study of US middle-aged adults found differential effects of two regulatory strategies, reappraisal and suppression, on CVD risk as assessed using the Framingham General CVD Risk Algorithm (167). Reappraisal has been identified as an adaptive strategy, whereas suppression can be adaptive in some situations but is more often found to be maladaptive (59). In this study, a one-standard-deviation increase in reappraisal scores were associated with 6.8% lower 10-year CVD risk, while a one standard deviation increase in suppression scores was associated with 11.6% higher risk. Associations were particularly robust for women as compared with men (167). Taken together, this emerging body of work strongly suggests that emotion regulation may contribute significantly to cardiovascular health.[6]

ACUTE EMOTIONS AND CHD

A separate set of mechanisms by which emotion may lead to CHD involves acute or "triggering" effects (for a more comprehensive review, see 86). For example, acute anxiety states may lead to hyperventilation, which then may trigger coronary vasospasm (168). It has also been hypothesized that acute hemodynamic stress caused by transient, intense emotional states may cause rupture of atherosclerotic plaques on the vessel wall of coronary arteries and initiate acute coronary events including sudden cardiac death (86). Triggering in this context is generally hypothesized to affect individuals who already have existing damage, and triggers have been defined as stimuli that cause acute pathophysiological changes that in turn lead to a cardiac event (169). A variety of studies have provided evidence that acute episodes of anger, anxiety, or depression (as well as stress) may serve as triggers for a cardiac event. For example, using a case-crossover design with 1,623 patients, the Determinants of Myocardial Infarction Onset Study reported evidence that episodes of anger and anxiety may be potent triggers of acute MI (170). The RR of MI in the 2 hours after an episode of anger was 2.3 (95% CI: 1.7 to 3.2), and after an episode of anxiety was 1.6 (95% CI: 1.1 to 2.2).

Investigators have also identified another form of acute myocardial dysfunction related to sudden emotional stress, which appears to occur in the absence of significant coronary disease. This syndrome has been variously referred to as acute myocardial stunning, broken heart syndrome, left ventricular apical ballooning syndrome, stress cardiomyopathy, and takotsubo cardiomyopathy. Since the publication of a landmark paper in the *New England Journal of Medicine* (171), this syndrome has received considerable attention, with a growing body of work providing additional evidence of this cardiotoxic effect of extreme emotion experience (172, 173). While some aspects are similar to the type of triggering described above, there are differences. Triggering effects occur in the context of plaque rupture, coronary thrombosis, and preexisting damage, and generally lead to irreversible damage; in contrast, stress cardiomyopathy occurs in the absence of obstructive coronary disease, is characterized by markedly elevated catecholamines and left ventricular

6 Emotion regulation also matters for health because it is an important component of self-regulation. The ability to regulate emotions, focus attention, resist temptation, and delay gratification are highly correlated and they all index prefrontal cortical activity (see Chapter 14). In turn, self-regulation indexes an individual's ability to make healthy behavioral choices over the lifecourse—for example, resist smoking, resist overeating, exercise regularly, brush their teeth, and so forth.

dysfunction, and has low risk of recurrence (174). Further evidence of triggering has been demonstrated in both men and women, whereas stress cardiomyopathy appears to occur more frequently among women. Emotional stress has been identified as a potential trigger for stress cardiomyopathy because it is often accompanied by dramatic elevations in catecholamines, one of the defining features of myocardial stunning. Though the precise pathophysiologic mechanisms have not yet been delineated, there is increasing consensus that this is a recognizable clinical syndrome occurring often in the context of severe emotional stress (174).

OTHER CARDIOMETABOLIC DISEASES

HYPERTENSION

Numerous studies have observed elevated blood pressure levels and higher prevalence of hypertension among individuals with anxiety or depressive disorders (e.g., 175, 176) or who report high levels of anxiety, depression, or anger (e.g., 177). While a number of earlier studies suggested high levels of negative emotions may be associated with increased risk of developing hypertension (e.g., 178), later studies often reported a null association (e.g., 179) or an inverse association (e.g., 180, 181). However, many of these studies are limited by methodologic problems (e.g., cross-sectional designs, low statistical power) and have focused on elderly adults (182). With a more systematic consideration of negative emotions in relation to incident hypertension using rigorous methods, findings may be different. One recent study considered trajectories of depression in relation to incident hypertension in middle-aged adults over 24 years of follow-up (182). Individuals with repeated depressive episodes compared with those with transient or persistently low levels of depression had a faster age-related increase in hypertension such that for every 5-year increase in age, there was a 7% greater increase in the odds of developing hypertension. The mixed findings in this literature have prompted some thought as to whether there may be circumstances whereby underlying pathophysiology alters both emotions and capacity to regulate blood pressure. For example, it may be that hypoperfusion and increased cerebral blood flow pulsatility lead to brain vascular damage that causes both depression and altered blood pressures (183). Thus, at present it appears that poorly regulated blood pressure often occurs with dysregulated emotions, but the direction of effects is as yet inconclusive. Future work will need to consider more carefully the possibility of bidirectional effects.

DIABETES

Type 2 diabetes (T2D) is a serious public health concern in its own right; it is a major risk factor for the development of CVD and is also characterized by many of the risk factors associated with CVD, including inflammation, neuroendocrine dysfunction, poor diet and sedentary lifestyle, and excessive alcohol consumption. Thus, T2D has increasingly been a focus for investigators interested in the relationship between emotions and health. Most work to date has investigated risk associated with depression. Three meta-analyses of the literature have been conducted, with all reporting a consistent positive association (184–186). The most recent meta-analysis summarized across 23 prospective studies that combined included 424,557 participants, had a mean follow-up of 8.3 years, and considered 19,977 cases of incident diabetes. Results indicated an increased risk of developing

diabetes among depressed versus nondepressed individuals, with an RR of 1.56 (95% CI: 1.37–1.77) that was somewhat attenuated after adjusting for conventional risk factors (adjusted RR = 1.38; 95% CI: 1.23–1.55). Findings from this latest meta-analysis are highly consistent with the earlier ones. Worth noting is that several of these meta-analyses have also found evidence for bidirectional effects, whereby having diabetes was associated with increased risk of developing depressive symptoms or disorder and vice versa (185, 187). While less work has considered anxiety or anger in relation to the onset of T2D, the small number of studies that have been conducted report suggestive findings with excess risks reported ranging from 35 to 50% (e.g., 188, 189). A recent review of the literature concluded that different forms of emotional distress may influence risk of developing T2D but that more rigorous research is needed to confirm the findings thus far (190).

METABOLIC SYNDROME AND OBESITY

Much of the work linking negative emotions with other cardiometabolic outcomes has focused primarily on depression. Directionality of effects are often debated, with evidence available that depression increases risk of developing metabolic syndrome and its component conditions like obesity or hypercholesterolemia but that these conditions may also lead to increased risk of experiencing depression (e.g., 191, 192). A small meta-analysis of four prospective studies with a total sample size of 3,834 participants reported a 52% (95% CI: 1.20–1.91) excess risk of developing metabolic syndrome associated with having depression at study baselines. Follow-up periods ranged from 6 to 17 years, and effects were stronger in women versus men (193). These findings were consistent with effect estimates derived from cross-sectional studies. Of note, when studies examining possible effects of metabolic syndrome on depression were meta-analyzed, a significant effect was also reported (RR = 1.49, 95% CI: 1.19–1.89), highlighting the bidirectional nature of effects (193).

Anxiety and anger are substantially less well studied in relation to metabolic syndrome and its component conditions. A review of the literature in 2007 identified only two prospective studies of anxiety and three of anger. Findings for anxiety were largely null, but positive associations were found more consistently with anger (194). The authors suggested that studies of anxiety were perhaps too limited to draw definitive conclusions and that bidirectional effects were also likely. Somewhat surprisingly, new studies in the area since publication of the review remain scarce.

Few studies have considered anger in relation to changes in weight status or developing obesity; more studies have considered risk of obesity associated with anxiety, suggesting some elevation in risk among individuals with high levels of anxiety (195). Recent work has focused more specifically on PTSD as a possible risk factor for weight gain (e.g., 196). Unlike other negative emotions, anxiety related to PTSD has a clear onset, which permits a more detailed examination of whether weight gain is simply more likely among those predisposed to be anxious or occurs in response to developing symptoms of anxiety. One recent study compared patterns of weight gain among women prior to and after they developed PTSD and reported more rapid weight gain in these women after PTSD onset, with no differences evident prior to PTSD onset when compared with women without PTSD (197). Relatedly, women with PTSD were also at significantly greater risk of becoming obese (OR = 1.36, 95% CI: 1.19–1.56).

CORONARY ATHEROSCLEROSIS

Cardiovascular disease is the end result of underlying atherosclerotic processes (198), and based on findings with CVD risk, a number of investigators have hypothesized that negative emotions may contribute directly to the progression of atherosclerosis (e.g., 199). Studies of atherosclerosis or related measures of subclinical CVD generally image some aspect of the vasculature to obtain an assessment of disease development prior to the emergence of frank CVD. Common subclinical CVD indices include carotid intima media thickness (IMT) and coronary artery calcification (CAC). A well-validated subclinical CVD indicator, CAC is identified using electron beam computed tomography that evaluates calcified plaques in the epicardial coronary system (200). Ultrasound assessment of the thickness of the lumen-intima interface and the media-adventitia interface of the carotid artery is used to assess IMT and obtain measures of carotid plaque.

Studies examining negative emotions like anxiety in relation to IMT and CAC have generally reported mixed findings (e.g., 201, 202). A recent review of anxiety and subclinical CVD markers concluded that studies thus far have shown a weak association (119) and the relationship with depression or anger appears similar (e.g., 203, 204). However, the majority of this literature is cross-sectional. Recent work has suggested that cross-sectional findings must be viewed with caution, as often even conventional CVD risk factors when measured concurrently do not predict some of these subclinical CVD outcomes as well as one would expect (205). In fact, findings from longitudinal studies have been somewhat stronger. For example, one study of 726 men and women showed that sustained high trait anxiety over time was related to marginally greater IMT progression over four years, and significantly higher plaque was evident at the follow-up assessment among men (206). One study of 324 men and women demonstrated greater 3-year change in carotid IMT associated with higher depressive symptoms (202); another study of 209 women also found higher levels of anger associated with greater changes in IMT over a 3-year period (207). Rozanski and colleagues (2011) suggest weak associations may be due in part to assessing the chronicity of emotion inadequately, to potential nonlinearities in the relationship with subclinical CVD markers, and also to inadequate assessment of the possibility of synergistic effects between multiple factors that contribute to the pathogenesis of atherosclerosis (205).

CANCER

Despite the long-held beliefs about a link between emotion and cancer, empirical support for the role of emotion in the development of cancer continues to be sparse. Based on work with cancer patients, Temoshok (208) proposed a model of the cancer-prone individual (Type C personality), as one who is stoic, has difficulty in expressing emotions, and has an attitude of resignation or helplessness/hopelessness. Empirical support for this model was reported in several studies (209, 210); however, subsequent work has not only identified serious methodological issues with these studies but also failed to replicate the findings (211, 212). To date, much of the work positing a relationship between emotion and cancer incidence has been based on the premise that effects are most likely mediated either through health behavior differences (e.g., smoking; 213) or through altered immune functioning (84, 214).

Because methodologically rigorous examination of the role of psychological factors in cancer *incidence* is challenging, true prospective studies are uncommon. However, extensive research has shown that psychosocial stress and related distress is associated with alterations in a range of processes that play an important role in carcinogenesis, including immunity, inflammation, insulin resistance, DNA repair capacity, cellular apoptosis, proliferation, angiogenesis, and telomere length. For example, stressed mice injected with ovarian cancer cells developed 2.5-fold larger tumors with more invasive phenotypes versus controls; 50% of stressed mice had metastases versus none in controls (215). Another study in mice reported that chronic stress led to increased tumor incidence and decreased tumor latency, suggesting a role in initiation (216). Studies investigating emotional processes directly in relation to biological processes known to be relevant in carcinogenesis (e.g., immune function and surveillance, DNA repair capacity, cellular apoptosis) may provide additional insight into whether emotions are involved in the etiology of cancer.

There are numerous studies of depression and cancer mortality that generally show that higher levels of depression are associated with elevated cancer mortality even after adjusting for potentially confounding medical variables (217). There is also a great deal of research on and evidence for the role of distress in the progression of and adjustment to cancer (e.g., 218). However, studies of cancer mortality or cancer progression cannot distinguish between the effects of emotion on cancer *incidence* versus *survival* following the diagnosis of cancer. The mechanisms linking emotion to cancer or survival may be specific to issues related to disease progression. For instance, disease progression may be related to psychosocial adjustment to the diagnosis and illness, adherence to treatment, and the availability of social support, including instrumental help with treatment regimens.

More recent studies of negative emotions and cancer incidence are suggestive but still limited, with most studies to date focused on depression as the risk factor. For example, one study considered depressive symptoms and prospective incidence of colorectal cancer among 81,612 initially healthy women followed over an 8-year period (219). Women with higher levels of depressive symptoms showed elevated risk of incident colorectal cancer after adjusting for a range of potential confounders (HR = 1.43, 95% CI: 0.9–2.11), and associations were strongest among overweight women. Several recent meta-analyses have considered the literature more systematically. One study reported an increased risk of incident cancer associated with depression when looking across multiple cancers, with effects stronger in studies with larger sample sizes ($n \geq 100{,}000$) and those conducted over a longer follow-up period (≥10 years) (220). Findings, however, varied depending on the specific cancer considered. This highlights one of the difficulties in studying cancer. Cancer comprises a heterogeneous group of diseases of multiple etiologies that vary in their tissue of origin, cell type, biological behavior, anatomic site, and degree of differentiation (221).

Another review summarized only studies of depression and breast cancer, and suggested that, while the overall meta-analysis did not find a strong relation with incident breast cancer, this may be due to several methodological problems (222). These include having an insufficient follow-up period, as the investigators suggest that a minimum of 18 years is necessary based on the amount of time it takes for tumor cells to develop into detectable tumors. Other methodological issues leading to heterogeneity in findings are the varied methods for assessing depression (e.g., number of somatic symptoms assessed) and the often limited number of case counts available in the studies. Thus, studies with an inappropriately short time frame and small case counts may tend to

underestimate associations, and such considerations should be included in future work in the area (222). Another issue in this area of research is that prebiopsy studies are sometimes considered as prospective studies. These studies investigate individuals who have been identified as needing diagnostic tests, and assess emotional functioning prior to obtaining the results of the tests (e.g., 223). However, in such situations individuals will have a self-prediction for what the outcome of their test will be; that prediction will almost certainly influence their responses to measures of distress. Thus, such studies have limited capacity to assess causality in the relation between emotional functioning and incident cancer (222). While some studies have considered the role of positive emotional functioning in relation to cancer progression or survival (224), almost no work has considered effects in relation to incident cancer. Moreover, empirical findings in the area have been mixed, and the idea that positive feeling and a "fighting spirit" might slow the progression and rate of disease development has been a source of significant controversy (e.g., 225, 226). Overall, direct evidence that psychological distress (or emotions more generally) is involved in the etiology of cancer in humans is limited; however, the known biologic consequences of recurring distress and their relevance to cancer initiation and promotion suggest that this issue deserves a closer and more rigorous examination.

INFECTIOUS OR IMMUNE-MEDIATED DISEASES AND OTHER HEALTH OUTCOMES

Negative emotions are thought to alter susceptibility to infectious diseases via effects on immune function (227). Psychological distress and stressors appear to be reliably associated with immune function down-regulation, although fewer studies have examined the effects of specific emotions (228, 229). For example, one recent study of healthy middle-aged men and women examined levels of psychological distress in relation to changes in markers of immune function over a 1-year period, including natural killer (NK), B, and T cell counts (230). The study found higher distress was associated with suppression of NK cell immunity (but not the other cell types) while initial NK cell counts did not predict subsequent levels of distress, suggesting that emotions may precede alterations in cellular immunity.

Whether findings with immune functioning may provide insight into likely effects of distress on immune-related hard health outcomes has been tested among healthy individuals in several ways. One line of work has used a viral challenge methodology in a controlled laboratory setting to examine the association between emotion (both positive and negative) and the common cold (227, 231). Given that only a proportion of people exposed to an infectious agent will develop clinical disease, investigators have examined whether emotions were associated with risk of developing a cold among individuals exposed to the virus. In these studies healthy subjects are exposed to a common cold virus, quarantined, and monitored for the development of biologically verified clinical illness (231, 232). Prior to exposure, levels of emotions are measured. Even after controlling for health behaviors, age, gender, and educational attainment, individuals with higher levels of negative emotions were more likely to develop clinical illness (232) while those expressing higher levels of positive emotions were at reduced risk (231, 233). Other work has found that psychological distress is associated with less robust immune response to vaccination (234) and an amplified inflammatory response after vaccination (235). Due to the methodological and

logistical complexities in conducting such studies, research in this area is not extensive and few studies have examined emotions other than general distress or depressive symptoms in relation to these outcomes.

Most research on emotion and other immune-mediated outcomes (e.g., HIV/AIDS, herpes simplex virus, asthma) has primarily considered disease progression, recurrence, or exacerbation, with less work assessing the role of emotion in the onset of disease. A limited number of studies have suggested that early-onset mental disorders (and specifically anxiety and depression) are associated with increased risk of adult-onset arthritis (236) and asthma (237). While these studies rely on retrospective reporting of onset of mental disorders, which may lead to an inflated association with any measured physical health outcome, one prospective study has provided consistent evidence. In this study, mental health problems were assessed among boys at the age of 8 who were followed through to adulthood, and findings suggested an increased risk of developing asthma among those who reported more depressive symptoms or other mental health problems in childhood (238). While findings on progression of immune-mediated diseases are suggestive and consistently indicate emotion is associated with exacerbations and recurrence (e.g., 239, 240, 241), they cannot directly evaluate the role of emotion in the development of these diseases.

A large body of work has repeatedly demonstrated that individuals with high levels of distress also have high levels of medically unexplained symptoms (242). Recurring sets of such medically unexplained symptoms have been characterized as "functional syndromes"—syndromes for which no pathologically defined changes in tissues can be found (e.g., fibromyalgia, chronic fatigue syndrome, irritable bowel syndrome, etc.) (243). Investigators have noted that these syndromes share many features, and speculate that this may be due to shared underlying similarities (242) potentially related to hyperexcitement of central neurons, referred to as central sensitization (244). Because physical symptoms frequently accompany affective experience, there is no consensus as to whether these symptoms should be viewed as primarily products of distress or as an important aspect of chronic illness.

The notion that emotional factors may play an important role in the development and exacerbation of these syndromes is commonly endorsed by patients and commonly proposed by investigators; however, empirical evidence lags in part due to the many methodological challenges that arise for this research. A particular challenge is related to the lack of "objective" indicators of these syndromes, which are largely diagnosed based on symptom reporting. A limited set of studies has considered whether traumatic or stressful events are likely to precede the development of these syndromes, with findings mixed or weak (245), but few studies have explicitly considered emotion per se. Other work has considered whether treatment for psychological distress might improve outcomes related to functional syndromes. One meta-analysis of treatment among fibromyalgia patients identified a small but consistent effect, whereby effective treatment of psychological distress reduced pain intensity and sleep problems (246). This area deserves more attention and, in fact, highlights another commonly noted phenomenon—that high levels of distress are generally accompanied by high levels of healthcare utilization (e.g., 247). Increasingly, this phenomenon appears to be creating an unmet public health challenge (248), further emphasizing the importance of understanding the interrelationships among affect, symptoms, and health.

NEW DIRECTIONS

The last decade has been marked by substantial progress. With these advances, the field is poised for new discoveries that have the potential to shift the paradigm for how public health and medical professionals view, target, and treat the interrelationships between mental and physical health. A number of technological, conceptual, and methodological innovations are in development that will facilitate generating powerful new knowledge. These areas are briefly described here.

A DEEPER UNDERSTANDING OF MECHANISMS

Research on emotion-health links has consistently suggested both biological and nonbiological pathways as jointly explaining how recurring emotion experience may cumulatively influence health outcomes. This work notwithstanding, there is continued skepticism by many in mainstream biomedicine with regard to the notion that mental states can directly influence health-related biology. This perspective was most clearly articulated by a past editor of the *New England Journal of Medicine*:

> In my 1985 editorial (249)…I wrote "the literature contains very few scientifically sound studies of the relation, if there is one, between mental state and disease."…What I was talking about was the view that mental state can *directly* cause or substantially modify organic disease independent of personal habits.…I am afraid my assessment of the literature has not changed very much in the 16 years since I wrote that. (250)

Such skeptics claim that any apparent causal relationship between emotions and health is spurious or a result of wishful thinking and that the research fails to make a convincing case. Two key critiques of the existing evidence on emotion and health have been offered: (1) the lack of plausible biological mechanisms and (2) methodological concerns. Multiple biological pathways have been proposed to explain how emotions might influence disease development. Countless studies have documented associations of emotions with biological mediators that are arguably linked with health, such as inflammatory biomarkers, heart rate variability, and levels of glucose control. Moreover, a great deal of work has considered links between stress-related physiological substrates (e.g., cortisol dysregulation, elevated catecholamines) and disease outcomes. However, such evidence may be considered inconclusive because studies that directly link emotion-specific neurobiological processes (e.g., specific neuropeptides or immune cells) with actual disease outcomes are scarce. Moreover, work that considers how emotions influence health at the cellular level has generally used logic to tie together studies on neural and molecular communications between the nervous and immune systems under conditions of stress, with epidemiological or experimental studies of emotion and health (e.g., 251, 252). Direct empirical evidence of biological alterations resulting from emotions and causing a disease outcome combined in a single study is rare and methodologically difficult to obtain. For example, anxiety is hypothesized to influence CHD development in part via chronically elevated inflammation profiles. However, studies that include measures of anxiety and alteration in levels of inflammation along with subsequent

development of disease have not been done. Perhaps as a result, the controversy around the premise that emotions directly change the course of disease development via some biological process continues (250).

On the other hand, emerging theories in behavioral economics point to the importance of emotions and emotional regulation in the adoption of health-related behaviors. Behavioral economists now recognize that acute emotional states (so-called anticipatory emotions) have a direct input into individuals' risk appraisals and behavioral decisions. In short, people often act "irrationally" (i.e., against their own long-term self-interest) while under the influence of "hot" states such as anger—for example, by experiencing a relapse of smoking after a heated argument with a spouse. Or depressed and anxious moods can lead to food cravings as a result of stimulation of the dopamine reward centers of the brain (253). The inability to regulate emotions can lead to a critical failure of self-control that can give rise to unhealthy behavioral patterns. In the search for "direct" effects of emotions on health outcomes, epidemiologists often statistically adjust for interindividual variations in health behaviors such as smoking, physical activity, and diet. In doing so, they may have "missed the gorilla in the room," and inappropriately controlled away one of the most interesting links between emotions and health.

That said, recent technological advances and greater interdisciplinary crosstalk suggest some exciting possibilities for work on potential biological mechanisms (for more detail see Chapter 14). This includes capacity for exploring relevant neurobiological substrates or cellular processes, and examining the role of social and emotional processes in regulating gene products involved in human pathogenic resistance and disease development. As multiple biological pathways are almost certainly involved, more comprehensive insight will be gained ultimately by considering these pathways in relation to each other.

For example, accelerated cellular aging, as measured by shortened telomere length and reduced telomerase activity in leukocytes, has been identified as a marker of increased risk for early morbidity and premature mortality across a range of disease outcomes (254, 255). Considering emotion in relation to speed of cellular aging may provide interesting insight into how or why emotions matter for health, particularly because studies could focus on either deteriorative processes (telomere shortening) or restorative processes (telomerase activity) or both. There is some work to suggest this may be a promising line of research. For example, one study found that individuals with high levels of emotion dysregulation as defined by chronic mood disorders had significantly shorter telomere length as compared with age-matched controls subjects without mood disorders. This finding was evident after adjusting for smoking and represents a difference reflecting approximately 10 years of accelerated aging (256).

Another potentially relevant biological substrate is oxytocin, a hormone involved in many social behaviors (257, 258). Recent theories have suggested oxytocin is part of a key neurobiological system underlying the link between positive social and emotional processes and health (259). Animal studies have clearly indicated that oxytocin is involved in numerous physiological functions (e.g., cardiovascular system, pancreas, kidney, brain). Research in humans has linked oxytocin to emotion and stress processes, noting its apparent stress-buffering effects (260, 261), though these have not been demonstrated to occur uniformly (e.g., 262, 263). Given the importance of social bonding and attachment for enabling individuals to develop effective emotion regulatory strategies, it may be informative to explore relationships between oxytocin activity, emotion processes, and health.

Capacity for characterizing the mechanisms through which genes and their products function and interact with each other and with the environment has also increased. Several recent studies have begun mapping the actions of genes and their biochemical pathways to psychological phenomena, suggesting that psychological experiences can modulate the expression of specific genes, the proteins they code for, and the physiological pathways they regulate (264, 265). For example, one study found highly versus less lonely individuals had increased expression of genes involved in immune activation, transcription control, and cell proliferation, and decreased expression of genes involved in innate antiviral resistance, supporting antibody production, and mature B lymphocyte function (266). Considering epigenetic mechanisms, processes that alter gene expression without actually changing the genetic code, as well as the interplay between emotions and the microbiome, may also provide insight into molecular pathways linking emotional processes with health over the lifecourse.

When considering biological or behavioral mechanisms underlying the relation between emotion and health, it is important to examine the full spectrum of effects. Effects of negative emotions are typically considered from the perspective of the harmful biological processes they set in motion, while effects of positive emotion are most often considered in relation to the harmful processes they *fail* to set in motion (e.g., reduced levels of inflammation). However, the relationship between emotion and health may also be a function of the absence or presence of restorative processes that are set into motion as well. This broader approach, recently described in some detail (40), refers not only to biological but also to behavioral mediators of health. For example, positive emotion may be associated with engaging in restorative health behaviors like regularly consuming fruits and vegetables and getting 8 hours of sleep each night, and also with less biological dysfunction like inflammation. Positive emotion may also directly reduce levels of experienced stress and/or buffer the harmful effects of stress (267). The notion that positive emotion may buffer effects of stress has received somewhat more empirical attention. For example, in several experiments, Fredrickson and colleagues demonstrated that cardiovascular reactivity following an initially stressful situation (e.g., preparing a speech) returned to baseline more quickly among individuals who experienced a positive mood than among individuals who experienced a neutral mood (268, 269).

A LIFECOURSE PERSPECTIVE

As may be evident from the review of the evidence, research to date has focused largely on emotion and health in adulthood. While informative, these studies may miss a myriad of processes already set in motion early in life. A lifecourse perspective can help to clarify how and why emotions might influence later health outcomes, and, as described earlier in the chapter, how the development of effective emotional functioning is influenced by the interaction of the larger social environment with the family environment. Recent epidemiologic and other work has begun to build compelling evidence that early childhood environments critically influence adult physical health (270). Repetti and colleagues (24) proposed a framework for how early family environments influence physical health and highlighted the critical nature of emotional processes that develop early in life. Emotions are fundamental to adaptation, providing both the impetus and mechanisms for adaptive behavior. The child development literature suggests that a major task of early childhood

is developing the capacity to regulate emotions (271). The ability to control frustration, delay gratification, or self-soothe contributes to or undermines the growth of new skills and competencies in young children that in turn matter for learning to evaluate risk and make decisions and for navigating social and other challenges. As discussed throughout the chapter, these factors are all strongly implicated in one's ability to maintain and protect health over the lifecourse. A lifecourse perspective sharpens the focus on emotion regulation or balancing positive and negative emotions, rather than considering single emotions in isolation. Moreover, it provides a conceptual framework for addressing questions about social factors that influence development of regulatory capacity, effects when regulatory capacity is not attained early in life, duration or timing (e.g., sensitive periods) of exposure, and reversibility of effects of emotion in relation to physical health.

Research on emotion regulation has been conducted both in the developmental and adult literatures, but integration between them is still limited, particularly regarding the relation with physical health (56). Regulating emotion is not simply learning to suppress emotions but is a broader skill that involves being able to employ and control emotions (both positive and negative) effectively in relationships and in a wide range of settings (271). Thus, emotion regulation involves changes in aspects of the emotion experience such as its magnitude, duration, or intensity. Gross and colleagues (56) have identified a constrained set of emotion regulatory processes and argue that particular forms of regulation are neither good nor bad, but rather their effectiveness depends on the context in which they are used. Regulation may occur via either intrinsic (by self) or extrinsic (by other) processes, but in childhood, extrinsic regulatory processes are more commonly employed.

Apart from the limited work considering emotion regulation in relation to cardiovascular outcomes described earlier, several other lines of work touch on this issue. Some research has been conducted on alexithymia, a condition characterized by the failure or inability to express emotion. This research has thus far suggested that it is a risk factor for symptoms and illness behavior but not for organic disease (272). Other work has examined whether disclosure of strong emotional feelings can improve health outcomes by avoiding the cumulative stress of inhibition. Positive health effects have been demonstrated (162), although several recent meta-analyses arrived at somewhat conflicting conclusions. The most recent meta-analysis found a significant and positive moderate effect of disclosure (273), but another summary of the research found no evidence of a clear effect (274). Differences between these studies may have been due to different selection criteria for the quantitative review. Generally, the findings thus far suggest that health will be optimized when emotion regulatory strategies facilitate acknowledging, expressing, and processing emotions appropriately (163).

Greater consideration of the role of emotion regulation in the maintenance of health or development of disease over the lifecourse is warranted. Dysregulation in childhood emotional functioning has been demonstrated to persist well into adulthood (e.g., 275, 276). These findings suggest that effects of emotions on adult disease processes may initiate much earlier in life than has generally been considered. For example, failure to learn appropriate strategies for emotion regulation in childhood may set up a lifelong pattern of yielding to temptation, immediate gratification ("myopia"), and the use of maladaptive behavioral coping responses to stress. Conversely, appropriate regulation may lead to accruing resilience. How early health-related risk (or resilience) associated with emotional functioning may be detected has not yet been determined but is important for informing the design of interventions that target where and when exposures have the highest

impact. The classic marshmallow test developed by Mischel suggests that variations in the ability to delay gratification are already apparent in 4-year-old preschoolers, and that lower delay discounting predicts both academic achievement as well as health behaviors (such as smoking, drug abuse, obesity) more than a decade later (277).

A challenge for research seeking to identify effects of emotion earlier in the lifecourse on later disease development (or health promotion) is the timeframe required for study, often spanning many years. One strategy is to identify and assess relevant biomarkers that may provide a measure of risk prior to developing disease. This also has the advantage of identifying individuals at high risk for whom greater surveillance and targeted interventions may significantly contribute to delaying disease onset. As a result, a growing body of work has begun to consider emotional functioning in relation to biomarkers associated primarily with increased risk of developing cardiometabolic disease. The strongest work to date has considered depression in relation to inflammatory markers, although there is some debate as to whether depression precedes inflammation or vice versa, or if effects are bidirectional. At present, it seems likely that effects are bidirectional; however, this work has also suggested that emotional distress, when it occurs earlier in life, may precede and trigger inflammatory processes. For example, one study reported that high levels of emotional distress assessed at age 7 years were associated with higher levels of CRP at age 42 years even after adjustment for relevant covariates including child health status (278). Another recent study found high levels of distress at age 8 years were associated with elevated levels of inflammation at age 10, but age 10 inflammatory levels were not associated with subsequent increases in distress (279).

BROADENING OUTCOMES AND METHODS

Problems with causal inference represent a major concern about the quality of the evidence for a causal relationship between emotion and disease development. Making the case that emotions influence health is bedeviled by the view that emotions are in fact irrational or superfluous. If emotions are simply epiphenomena, it is difficult to argue that they have material and lasting consequences. Moreover, much of the strongest evidence for the relationship is derived from longitudinal observational cohort studies that rely on self-report of emotion and may be susceptible to concerns about unmeasured confounders. To avoid concerns about self-report bias, studies have emphasized outcomes that can be determined objectively and for which there is a relatively clear onset, leading to a preponderance of evidence in the domain of cardiovascular disease. While emotions are almost certainly involved in the development of other outcomes such as cancer, autoimmune disease, or pain syndromes, because these outcomes are often diagnosed by self-reports of symptoms or have long latency periods, studies of these outcomes are vulnerable to even more criticism and skepticism.

Moreover, because of the observational nature of even studies with objectively measured outcomes that have a relatively clear onset, opponents of the hypothesis can argue that causality runs from disease to emotion, or that there is some underlying factor (e.g., genetics, low SES) driving both emotion and disease states. As a result, even the best studies continue to be criticized by some as too flawed to provide objective and credible evidence (e.g., 280). Because randomized clinical trials are generally considered the gold standard for evidence, inconsistent findings on

whether psychosocial interventions improve physical health outcomes in randomized controlled trials (within patient populations) have further fueled this concern (see Chapter 11, this volume). However, most trials have been conducted within cardiac patient populations rather than on initially healthy individuals. As a result, effects of distress are considered among individuals where significant biological damage has already been incurred, and it is unclear whether mitigating distress at this late stage could limit further damage or prevent further effects of existing damage. In addition, these kinds of trials are typically conducted among individuals in middle or later adulthood, but the duration, intensity, or reversibility of exposure to recurring emotional states has not been established.

A lifecourse perspective suggests that emotional processes are laid down early and, as a result, by middle age individuals have long been exposed to recurring emotional states. This is supported by studies reporting that among patients who have had a heart attack, a significant fraction report having a lifetime history of depression or anxiety occurring prior to the occurrence of their heart attack (281). Innovative study designs are needed to allow the field to make the case more convincingly and to broaden the outcomes under inquiry. A key question that is amenable to empirical inquiry (but has not yet been considered) is whether successful mitigation of emotional distress or setting up positive states within a healthy population effectively reduces risk of developing heart disease or other health conditions.

A number of other strategies are currently available that may help reduce concerns about causal inference in this research. Longitudinal studies that assess emotional functioning earlier in the lifecourse (e.g., in childhood) can help to reduce concerns about reverse causality by more convincingly establishing the temporal ordering in the relationship. Instrumental variable analysis and Mendelian randomization methods can be used to simulate an experiment and assess effects of emotions in observational data using a genotype that influences the emotion phenotype (e.g., anxiety) to estimate the effect of that emotion on a health outcome. Such studies might provide powerful evidence but are presently hampered by the failure to find strong evidence of either single genes or genetic scores that clearly underlie each emotion (for a genetic instrument) (282). Marginal structural modeling is another technique increasingly used to mitigate another major threat to causal inference, the possibility of "time-varying-confounding." This refers to the likelihood that the behaviors and risk factors that are subsequently affected by exposure to a given emotion may also increase risk of having that emotion. For example, anxiety may increase risk of being sedentary, but being sedentary may also increase likelihood of being anxious. Marginal structural models can be used to account for the potential for dynamic feedback processes by which behavioral and biological changes can act as either confounders or mediators of the effects of emotions on health outcomes.

CONCLUSIONS

Research increasingly supports the role of emotions in pathogenesis or disease protection, with findings strongest in relation to coronary heart disease. In this chapter, we have considered the state of the evidence to date. We have also highlighted future directions that will yield greater understanding of the mechanisms by which the experience of emotions—both positive and

negative—leads to differential health outcomes. This work will also build a stronger foundation for developing effective prevention and intervention strategies toward improving population health. Continued research on emotions is critical to the advancement of social epidemiology for at least three reasons: first, because the social patterning of emotions offers researchers an important clue about how variations in the external social environment produce differences in individual health status; second, because emotions represent a crucial link in the chain of causation that runs from stressors to biological responses within individuals (the so-called sociobiological translation); and third, because research on emotions provides a basis for the development of psychosocial interventions that aim to break or improve the link between social conditions and health outcomes.

REFERENCES

1. Plato. [Dialogues. English and Greek] Charmides; Alcibiades I and II; Hipparchus; The lovers; Theages; Minos; Epinomis/ Plato. Cambridge, MA: Harvard University Press; 1927.
2. Taylor SE, Repetti R, Seeman T. What is an unhealthy environment and how does it get under the skin? Annu Rev Psychol. 1997;48:411–47.
3. Williams SJ, Bendelow G. Emotions, health and illness: the "missing link" in sociology? In: James V, Gabe J, editors. Health and the sociology of emotions. Cambridge, MA: Blackwell Publishers; 1996. pp. 25–54.
4. Matthews KA, Gallo LC, Taylor SE. Are psychosocial factors mediators of socioeconomic status and health connections? A progress report and blueprint for the future. Ann N Y Acad Sci. 2010;1186:146–73.
5. DeSteno D, Gross JJ, Kubzansky L. Affective science and health: the importance of emotion and emotion regulation. Health Psychol. 2013;32(5):474–86.
6. Kemper TD. Sociological models in the explanation of emotions. In: Lewis M, Haviland JM, editors. Handbook of emotions. New York: Guilford Press; 1993. pp. 41–52.
7. Lazarus RS. Target article: Theory-based stress measurement. Psychol Inq. 1990;1(1):3–13.
8. Pearlin L, Lieberman M, Menaghan E, Mullen JT. The stress process. J Health Soc Behav. 1981;22:337–56.
9. Kessler RC. A disaggregation of the relationship between socioeconomic status and psychological distress. Am Sociol Rev. 1982;47(6):752–64.
10. Melchior M, Berkman LF, Niedhammer I, Zins M, Goldberg M. The mental health effects of multiple work and family demands: a prospective study of psychiatric sickness absence in the French GAZEL study. Soc Psychiatry Psychiatr Epidemiol. 2007;42(7):573–82.
11. Pearlin LI, Lieberman MA. Social sources of emotional distress. In: Simmons R, editor. Research in community and mental health. Greenwich, CT: JAI Press; 1978.
12. Turner RJ, Lloyd DA. The stress process and the social distribution of depression. J Health Soc Behav. 1999;40:374–404.
13. Koolhaas JM, Bartolomucci A, Buwalda B, de Boer SF, Flugge G, Korte SM, et al. Stress revisited: a critical evaluation of the stress concept. Neurosci Biobehav Rev. 2011;35(5):1291–301.
14. Yehuda R, LeDoux J. Response variation following trauma: a translational neuroscience approach to understanding PTSD. Neuron. 2007;56(1):19–32.
15. Marmot M, Wilkinson RG. Psychosocial and material pathways in the relation between income and health: a response to Lynch et al. BMJ. 2001;322(7296):1233–6.
16. Mendelson T, Thurston RC, Kubzansky LD. Affective and cardiovascular effects of experimentally-induced social status. Health Psychol. 2008;27(4):482–9.

17. Lazarus RS. Progress on a cognitive-motivational-relational theory of emotion. Am Psychol. 1991;46(8): 819–34.
18. McLeod JD, Kessler RC. Socioeconomic status differences in vulnerability to undesirable life events. J Health Soc Behav. 1990;31:162–72.
19. Chen E, Matthews KA. Cognitive appraisal biases: an approach to understanding the relation between socioeconomic status and cardiovascular reactivity in children. Ann Behav Med. 2001;23(2):101–11.
20. Matthews KA, Raikkonen K, Everson SA, Flory JD, Marco CA, Owens JF, et al. Do the daily experiences of healthy men and women vary according to occupational prestige and work strain? Psychosom Med. 2000;62(3):346–53.
21. Kessler RC, Neighbors HW. A new perspective on the relationships among race, social class, and psychological distress. J Health Soc Behav. 1986;27(June):107–15.
22. Ulbrich PM, Warheit GJ, Zimmerman RS. Race, socioeconomic status, and psychological distress: an examination of differential vulnerability. J Health Soc Behav. 1989;30:131–46.
23. Cochran SD, Mays VM, Sullivan JG. Prevalence of mental disorders, psychological distress, and mental health services use among lesbian, gay, and bisexual adults in the United States. J Consult Clin Psychol. 2003;71(1):53–61.
24. Repetti RL, Taylor SE, Seeman TE. Risky families: family social environments and the mental and physical health of the offspring. Psychol Bull. 2002;128(2):330–6.
25. Evans GW, Gonnella C, Marcynyszyn LA, Gentile L, Salpekar N. The role of chaos in poverty and children's socioemotional adjustment. Psychol Sci. 2005;16(7):560–5.
26. John OP, Gross JJ. Healthy and unhealthy emotion regulation: personality processes, individual differences, and life span development. J Pers. 2004;72(6):1301–33.
27. Gottman JM, Katz LF, Hooven C. Parental meta-emotion philosophy and the emotional life of families: theoretical models and preliminary data. J Fam Psychol. 1996;10(3):243–68.
28. Matthews KA, Gallo LC. Psychological perspectives on pathways linking socioeconomic status and physical health. Annu Rev Psychol. 2011;62:501–30.
29. Siegman AW, Smith TW. Introduction. In: Siegman AW, Smith TW, editors. Anger, hostility, and the heart. Hillsdale, NJ: Erlbaum; 1994. pp. vii–xv.
30. Allport GW. Pattern and growth in personality. New York: Holt, Rinehart, & Winston; 1961.
31. Harvey W. Exercitatio anetomica de motu cordis et sanguinis [An anatomical exercise concerning the movement of heart and blood]. Facsimile of original 1628 edition ed. London: Baillieve, Tindall, & Cox; 1928.
32. Alexander FG, French TM, Pollack GH. Psychosomatic specificity: experimental study and results. Chicago: The University of Chicago Press; 1968.
33. Siegman AW. From Type A to hostility to anger: reflections on the history of coronary-prone behavior. In: Siegman W, Smith TW, editors. Anger, hostility and the heart. Hillsdale, NJ: Erlbaum; 1994. pp. 1–21.
34. Friedman HS, Booth-Kewley S. The "disease-prone personality": a meta-analytic view of the construct. Am Psychol. 1987;42(6):539–55.
35. Elderon L, Whooley MA. Depression and cardiovascular disease. Prog Cardiovasc Dis. 2013;55(6):511–23.
36. Whooley MA, Wong JM. Depression and cardiovascular disorders. Ann Rev Clin Psych. 2013;9:327–54.
37. Edmondson D, Richardson S, Falzon L, Davidson KW, Mills MA, Neria Y. Posttraumatic stress disorder prevalence and risk of recurrence in acute coronary syndrome patients: a meta-analytic review. PLoS One. 2012;7(6):e38915.
38. Macleod J, Davey Smith G. Psychosocial factors and public health: a suitable case for treatment? J Epidemiol Commun H. 2003;57(8):565–70.

39. Everson-Rose SA, Lewis TT. Psychosocial factors and cardiovascular diseases. Annu Rev Public Health. 2005;26:469–500.
40. Boehm JK, Kubzansky LD. The heart's content: the association between positive psychological well-being and cardiovascular health. Psychol Bull. 2012;138(4):655–91.
41. Kobau R, Seligman ME, Peterson C, Diener E, Zack MM, Chapman D, et al. Mental health promotion in public health: perspectives and strategies from positive psychology. Am J Public Health. 2011;101(8): e1–9.
42. Barlow DH. Anxiety and its disorders. New York: Guilford Press; 1988.
43. Frijda NH. The emotions. Cambridge, UK: Cambridge University Press; 1986.
44. Arnold MB. Emotion and personality. New York: Columbia University Press; 1960.
45. Lazarus RS, editor. Emotions and adaptation: conceptual and empirical relations. Lincoln: University of Nebraska Press; 1968.
46. Frijda NH. Emotions are functional, most of the time. In: Ekman P, Davidson RJ, editors. The nature of emotion. New York: Oxford University Press; 1994. pp. 112–22.
47. Scherer KR. Emotion as a process: function, origin and regulation. Soc Sci Inform. 1982;21(4–5):555–70.
48. Fredrickson BL. What good are positive emotions? Rev Gen Psychol. 1998;2(3):300–19.
49. Lazarus R. The stable and the unstable in emotion. In: Ekman P, Davidson RJ, editors. The nature of emotion. New York: Oxford University Press; 1994. pp. 70–85.
50. Frijda NH. Moods, emotions episodes, and emotions. In: Lewis M, Haviland JM, editors. Handbook of emotions. New York: Guilford Press; 1993. pp. 381–405.
51. Breckler SJ. Emotion and attitude change. In: Lewis M, Haviland JM, editors. Handbook of emotions. New York: Guilford Press; 1993. pp. 461–74.
52. Fishbein M, Ajzen I. Belief, attitude, intention, and behavior: an introduction to theory and research. Reading, MA: Addison-Wesley; 1975.
53. Scheier MF, Bridges MW. Person variables and health: personality predispositions and acute psychological states as shared determinants for disease. Psychosom Med. 1995;57:255–68.
54. Smith TW, Spiro A. Personality, health, and aging: prolegomenon for the next generation. J Pers. 2002;36: 363–94.
55. Salovey P, Rothman AJ, Detweiler JB, Steward WT. Emotional states and physical health. Am Psychol. 2000;55(1):110–21.
56. Gross JJ, Thompson R. Emotion regulation: conceptual foundation. In: Gross JJ, editor. Handbook of emotion regulation. New York: Guilford; 2007.
57. Shedler J, Mayman M, Manis M. The illusion of mental health. Am Psychol. 1993;48(11):1117–31.
58. Pennebaker J, Beall SK. Confronting a traumatic event: toward an understanding of inhibition and disease. J Abnorm Psychol. 1986;95:274–81.
59. Gross JJ, John OP. Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. J Pers Soc Psychol. 2003;85(2):348–62.
60. Gross JJ. Emotion regulation in adulthood: timing is everything. Curr Dir Psychol Sci. 2001;10(6): 214–9.
61. Appleton AA, Kubzansky LD. Emotion regulation and cardiovascular disease risk. In: Gross JJ, editor. Handbook of emotion regulation. 2nd ed. New York: Guilford; 2013.
62. Kawachi I, Sparrow D, Vokonas PS, Weiss ST. Decreased heart rate variability in men with phobic anxiety (data from the Normative Aging Study). Am J Cardiol. 1995;75(14):882–5.
63. Baum A, Grunberg N. Measurement of stress hormones. In: Cohen S, Kessler RC, Gordon LU, editors. Measuring stress: a guide for health and social scientists. New York: Oxford University Press; 1995. pp. 175–93.

64. Ader R, Felten DL, Cohen N, editors. Psychoneuroimmunology. San Diego, CA: Academic Press; 1991.

65. Rabin BS, Cohen S, Ganguli R, Lysle DR, Cunnick JE. Bidirectional interaction between the central nervous system and the immune system. Crit Rev Immunol. 1989;9:279–312.

66. Deverts DJ, Cohen S, DiLillo VG, Lewis CE, Kiefe C, Whooley M, et al. Depressive symptoms, race, and circulating C-reactive protein: the Coronary Artery Risk Development in Young Adults (CARDIA) study. Psychosom Med. 2010;72(8):734–41.

67. Miller GE, Rohleder N, Stetler C, Kirschbaum C. Clinical depression and regulation of the inflammatory response during acute stress. Psychosom Med. 2005;67(5):679–87.

68. Boehm JK, Williams DR, Rimm EB, Ryff C, Kubzansky LD. Relation between optimism and lipids in midlife. Am J Cardiol. 2013.

69. Steptoe A, Demakakos P, de Oliveira C, Wardle J. Distinctive biological correlates of positive psychological well-being in older men and women. Psychosom Med. 2012;74(5):501–8.

70. Kok BE, Fredrickson BL. Upward spirals of the heart: autonomic flexibility, as indexed by vagal tone, reciprocally and prospectively predicts positive emotions and social connectedness. Biol Psychol. 2010;85(3):432–6.

71. Everson SA, Kauhanen J, Kaplan GA, Goldberg DE, Julkunen J, Tuomilehto J, et al. Hostility and increased risk of mortality and acute myocardial infarction: the mediating role of behavioral risk factors. Am J Epidemiol. 1997;146:142–52.

72. Brody AL, Olmstead RE, Abrams AL, Costello MR, Khan A, Kozman D, et al. Effect of a history of major depressive disorder on smoking-induced dopamine release. Biol Psychiatry. 2009;66(9):898–901.

73. Cohen S, Rodriguez MS. Pathways linking affective disturbances and physical disorders. Health Psychol. 1995;14:374–80.

74. Baumeister RF, Tierney J. Willpower: rediscovering the greatest human strength. New York: Penguin Press; 2011.

75. Ingjaldsson JT, Laberg JC, Thayer JF. Reduced heart rate variability in chronic alcohol abuse: relationship with negative mood, chronic thought suppression, and compulsive drinking. Biol Psychiatry. 2003;54(12):1427–36.

76. Garland EL, Franken IH, Howard MO. Cue-elicited heart rate variability and attentional bias predict alcohol relapse following treatment. Psychopharmacology (Berl). 2012;222(1):17–26.

77. McGonigal K. The willpower instinct: how self-control works, why it matters, and what you can do to get more of it. New York: Penguin; 2012.

78. Krygier JR, Heathers JA, Shahrestani S, Abbott M, Gross JJ, Kemp AH. Mindfulness meditation, well-being, and heart rate variability: a preliminary investigation into the impact of intensive Vipassana meditation. Int J Psychophysiol. 2013;89(3):305–13.

79. Mullainathan S, Shafir E. Scarcity: why having too little means so much. New York: Times Books; 2013.

80. Betensky JD, Contrada RJ. Depressive symptoms, trait aggression, and cardiovascular reactivity to a laboratory stressor. Ann Behav Med. 2010;39(2):184–91.

81. Treiber FA, Kamarck TW, Schneiderman N, Sheffield D, Kapuku G, Taylor T. Cardiovascular reactivity and development of preclinical and clinical disease states. Psychosom Med. 2003;65:46–62.

82. Chida Y, Steptoe A. Greater cardiovascular responses to laboratory mental stress are associated with poor subsequent cardiovascular risk status: a meta-analysis of prospective evidence. Hypertension. 2010;55(4):1026–32.

83. Carroll D, Ginty AT, Der G, Hunt K, Benzeval M, Phillips AC. Increased blood pressure reactions to acute mental stress are associated with 16-year cardiovascular disease mortality. Psychophysiology. 2012;49(10):1444–8.

84. Segerstrom SC, Miller GE. Psychological stress and the human immune system: a meta-analytic study of 30 years of inquiry. Psychol Bull. 2004;130(4):601–30.
85. Steptoe A, Hamer M, Chida Y. The effects of acute psychological stress on circulating inflammatory factors in humans: a review and meta-analysis. Brain Behav Immun. 2007;21(7):901–12.
86. Bhattacharyya MR, Steptoe A. Emotional triggers of acute coronary syndromes: strength of evidence, biological processes, and clinical implications. Prog Cardiovasc Dis. 2007;49(5):353–65.
87. Mittleman MA, Maclure M, Nachnani M, Sherwood JB, Muller JE. Educational attainment, anger, and the risk of triggering myocardial infarction onset. Arch Intern Med. 1997;157(7):769–75.
88. Steptoe A, Strike PC, Perkins-Porras L, McEwan JR, Whitehead DL. Acute depressed mood as a trigger of acute coronary syndromes. Biol Psychiatry. 2006;60(8):837–42.
89. Selye H. The physiology and pathology of exposure to stress. Montreal: Acta; 1950.
90. DeLongis A, Coyne JC, Dakof G, Folkman S, Lazarus RS. Relationship of daily hassles, uplifts, and major life events to health status. Health Psychol. 1982;1:119–36.
91. Dohrenwend BS, Dohrenwend BP. Stressful life events: their nature and effects. New York: Wiley; 1974.
92. Selye H. Confusion and controversy in the stress field. J Human Stress. 1975;1(2):37–44.
93. Lazarus RS. From psychological stress to the emotions: a history of changing outlooks. Annu Rev Psychol. 1993;44:1–21.
94. Cohen S, Kessler RC, Gordon LU. Strategies for measuring stress in studies of psychiatric and physical disorders. In: Cohen S, Kessler RC, Gordon LU, editors. Measuring stress: a guide for health and social scientists. New York: Oxford University Press; 1995.
95. Spielberger CD, Sarason JG. Stress and anxiety. Washington, DC: Hemisphere; 1978.
96. Lader M, Marks I. Clinical anxiety. London: Heinemann; 1973.
97. Ohman A. Fear and anxiety as emotional phenomena: clinical phenomenology, evolutionary perspectives, and information-processing mechanisms. In: Lewis M, Haviland JM, editors. Handbook of emotions. New York: Guilford Press; 1993. pp. 511–36.
98. Barlow DH. Unraveling the mysteries of anxiety and its disorders from the perspective of emotion theory. Am Psychol. 2000;55(11):1247–63.
99. Kubzansky LD, Kawachi I. Going to the heart of the matter: Do negative emotions cause coronary heart disease? J Psychosom Res. 2000;48:323–37.
100. Izard C. Human emotions. New York: Plenum; 1977.
101. Tomkins S. Affect, imagery, and consciousness. Vol. 11: The negative affects. New York: Springer; 1963.
102. Stone AA. Measurement of affective response. In: Cohen S, Kessler RC, Gordon LU, editors. Measuring stress. New York: Oxford University Press; 1995. pp. 148–71.
103. Barefoot JC, Lipkus IM. The assessment of anger and hostility. In: Siegman AW, Smith TW, editors. Anger, hostility and the heart. Hillsdale, NJ: Erlbaum; 1994.
104. Robinson JP, Shaver PR, Wrightsman LS, editors. Measures of personality and social psychological attitudes. New York: Academic Press; 1991.
105. Lazarus RS. Emotion and adaptation. New York: Oxford University Press; 1991.
106. Leventhal H, Patrick-Miller L. Emotions and physical illness: Causes and indicators of vulnerability. In: Lewis M, Haviland-Jones JM, editors. Handbook of emotions. 2nd ed. New York: Guilford Press; 2000. pp. 523–37.
107. Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. BMC Med. 2013;11:126.
108. Ekman P, Levenson RW, Friesen WV. Autonomic nervous system activity distinguishes among emotions. Science. 1983;221:1208–10.

109. LeDoux JE. Emotion-specific physiological activity: Don't forget about CNS physiology. In: Ekman P, Davidson RJ, editors. The nature of emotions. New York: Oxford University Press; 1994. pp. 248–52.

110. Schwartz GE, Weinberger DA. Patterns of emotional responses to affective situations: relations among happiness, sadness, anger, fear, depression, and anxiety. Motiv Emotion. 1980;4:175–91.

111. Smith CA, Ellsworth PC. Patterns of appraisal and emotion in taking an exam. J Pers Soc Psychol. 1987;52(3):1–14.

112. Tomaka J, Blascovich J, Kelsey RM, Leitten CL. Subjective, physiological, and behavioral effects of threat and challenge appraisal. J Pers Soc Psychol. 1993;65(2):248–60.

113. Clark LA, Watson D. Tripartite model of anxiety and depression: psychometric considerations and taxonomic implications. J Abnorm Psychol. 1991;100:316–36.

114. Lerner J, Keltner D. Beyond valence: Toward a model of emotion-specific influences on judgment and choice. Cogn Emot. 2000;14(4):473–93.

115. Fjeldsoe B, Neuhaus M, Winkler E, Eakin E. Systematic review of maintenance of behavior change following physical activity and dietary interventions. Health Psychol. 2011;30(1):99–109.

116. Clark LA. The anxiety and depressive disorders: Descriptive psychopathology and differential diagnosis. In: Kendall PC, Watson D, editors. Anxiety and depression: distinctive and overlapping features. New York: Academic Press; 1989. pp. 83–129.

117. Kubzansky LD. Key 2010 publications in behavioral medicine. Cleve Clin J Med. 2011;7 (Suppl 1):S65–8.

118. Phillips AC, Batty GD, Gale CR, Deary IJ, Osborn D, MacIntyre K, et al. Generalized anxiety disorder, major depressive disorder, and their comorbidity as predictors of all-cause and cardiovascular mortality: the Vietnam experience study. Psychosom Med. 2009;71(4):395–403.

119. Thurston RC, Rewak M, Kubzansky LD. An anxious heart: anxiety and the onset of cardiovascular diseases. Prog Cardiovasc Dis. 2013;55(6):524–37.

120. Mathers CD, Lopez AD, Murray CJL. The burden of disease and mortality by condition: data, methods, and results for 2001. In: Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJL, editors. Global burden of disease and risk factors. Washington, DC: World Bank; 2006.

121. World Health Organization. Global status report on noncommunicable diseases 2010. Geneva: WHO; 2011.

122. Beaglehole R, Magnus P. The search for new risk factors for coronary heart disease: occupational therapy for epidemiologists? Int J Epidemiol. 2002;31(6):1117–22; author reply 34–5.

123. Lloyd-Jones DM, Hong Y, Labarthe D, Mozaffarian D, Appel LJ, Van Horn L, et al. Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic Impact Goal through 2020 and beyond. Circulation. 2010;121(4):586–613.

124. Strasser T. Reflections on cardiovascular diseases. Interdiscip Sci Rev. 1978;3:225–30.

125. Friedman M, Rosenman RH. Association of specific overt behavior pattern with blood and cardiovascular findings. JAMA. 1959;169:1286–96.

126. Review Panel on Coronary-Prone Behavior and Coronary Heart Disease. Coronary-prone behavior and coronary heart disease: a critical review. Circulation. 1981;63:1169–215.

127. Allan R, Scheidt S. Stress, anger, and psychosocial factors for coronary heart disease. In: Manson JE, Ridker PM, Gaziano JM, Hennekens CH, editors. Prevention of myocardial infarction. New York: Oxford University Press; 1996.

128. Matthews KA. Coronary heart disease and Type A behaviors: update on and alternative to the Booth-Kewley and Friedman (1987) quantitative review. Psychol Bull. 1988;104:373–80.

129. Kawachi I, Sparrow D, Kubzansky LD, Spiro A, 3rd, Vokonas PS, Weiss ST. Prospective study of a self-report type A scale and risk of coronary heart disease: test of the MMPI-2 type A scale. Circulation. 1998;98(5):405–12.

130. Nunes EV, Frank KA, Kornfield DS. Psychologic treatment for Type A behavior pattern and for coronary heart disease: a meta-analysis of the literature. Psychosom Med. 1987;48:159–73.

131. Suls J, Bunde J. Anger, anxiety, and depression as risk factors for cardiovascular disease: the problems and implications of overlapping affective dispositions. Psychol Bull. 2005;131(2):260–300.

132. Booth-Kewley S, Friedman HS. Psychological predictors of heart disease: a quantitative review. Psychol Bull. 1987;101:343–62.

133. Rugulies R. Depression as a predictor for coronary heart disease. Am J Prev Med. 2002;23(1):51–61.

134. Wulsin LR, Singal BM. Do depressive symptoms increase the risk for the onset of coronary disease? A systematic quantitative review. Psychosom Med. 2003;65(2):201–10.

135. Roest AM, Martens EJ, de Jonge P, Denollet J. Anxiety and risk of incident coronary heart disease: a meta-analysis. J Am Coll Cardiol. 2010;56(1):38–46.

136. Chida Y, Steptoe A. The association of anger and hostility with future coronary heart disease: a meta-analytic review of prospective evidence. J Am Coll Cardiol. 2009;53(11):936–46.

137. Kubzansky LD. Sick at heart: the pathophysiology of negative emotions. Cleve Clin J Med. 2007;74(Suppl 1):S67–S72.

138. Smith TW. Hostility and health: Current status of a psychosomatic hypothesis. Health Psychol. 1992;11:139–50.

139. Matthews KA. Assessment of Type A behavior, anger, and hostility in epidemiological studies of cardiovascular disease. In: Ostfeld AM, Eaker ED, editors. Measuring psychosocial variables in epidemiologic studies of cardiovascular disease. Washington, DC: US Department of Health and Human Services, National Institutes of Health; 1985. pp. 153–84.

140. Miller TQ, Smith TW, Turner CW, Guijarro ML, Hallet AJ. A meta-analytic review of research on hostility and physical health. Psychol Bull. 1996;119(2):322–48.

141. Kawachi I, Sparrow D, Spiro A, Vokonas P, Weiss ST. A prospective study of anger and coronary heart disease: the Normative Aging Study. Circulation. 1996;94(9):2090–5.

142. Williams JE, Nieto FJ, Sanford CP, Tyroler HA. Effects of an angry temperament on coronary heart disease risk: the Atherosclerosis Risk in Communities Study. Am J Epidemiol. 2001;154(3):230–5.

143. Barlow DH. Disorders of emotions: clarification, elaboration, and future directions. Psychol Inq. 1991;2(1):97–105.

144. Haines AP, Imeson JD, Meade TW. Phobic anxiety and ischaemic heart disease. BMJ (Clinical Research Ed.). 1987;295(6593):297–9.

145. Smoller JW, Pollack MH, Wassertheil-Smoller S, Jackson RD, Oberman A, Wong ND, et al. Panic attacks and risk of incident cardiovascular events among postmenopausal women in the Women's Health Initiative Observational Study. Arch Gen Psychiatry. 2007;64(10):1153–60.

146. Dedert EA, Calhoun PS, Watkins LL, Sherwood A, Beckham JC. Posttraumatic stress disorder, cardiovascular, and metabolic disease: a review of the evidence. Ann Behav Med. 2010;39(1):61–78.

147. Jordan HT, Miller-Archie SA, Cone JE, Morabia A, Stellman SD. Heart disease among adults exposed to the September 11, 2001 World Trade Center disaster: results from the World Trade Center Health Registry. Prev Med. 2011;53(6):370–6.

148. Player MS, Peterson LE. Anxiety disorders, hypertension, and cardiovascular risk: a review. Int J Psychiatry Med. 2011;41(4):365–77.

149. Anda R, Williamson D, Jones D, Macea C, Eaker E, Glassman A, et al. Depressed affect, hopelessness, and the risk of ischemic heart disease in a cohort of U.S. adults. Epidemiology. 1993;4:285–94.

150. Van der Kooy K, van Hout H, Marwijk H, Marten H, Stehouwer C, Beekman A. Depression and the risk for cardiovascular diseases: systematic review and meta analysis. Int J Geriatr Psychiatry. 2007;22(7):613–26.

151. Davidson KW, Schwartz JE, Kirkland SA, Mostofsky E, Fink D, Guernsey D, et al. Relation of inflammation to depression and incident coronary heart disease (from the Canadian Nova Scotia Health Survey [NSHS95] Prospective Population Study). Am J Cardiol. 2009;103(6):755–61.

152. Whang W, Kubzansky LD, Kawachi I, Rexrode KM, Kroenke CH, Glynn RJ, et al. Depression and risk of sudden cardiac death and coronary heart disease in women: results from the Nurses' Health Study. J Am Coll Cardiol. 2009;53(11):950–8.

153. Pan A, Sun Q, Okereke OI, Rexrode KM, Hu FB. Depression and risk of stroke morbidity and mortality: a meta-analysis and systematic review. JAMA. 2011;306(11):1241–9.

154. Pan A, Okereke OI, Sun Q, Logroscino G, Manson JE, Willett WC, et al. Depression and incident stroke in women. Stroke. 2011;42(10):2770–5.

155. Kohler S, Verhey F, Weyerer S, Wiese B, Heser K, Wagner M, et al. Depression, non-fatal stroke and all-cause mortality in old age: a prospective cohort study of primary care patients. J Affect Disord. 2013;150(1):63–9.

156. Narayan, SM, Stein, MB. Do depression or antidepressants increase cardiovascular mortality? Journal of the American College of Cardiology. 2009;53(11):959–61.

157. Hamer M, David Batty G, Seldenrijk A, Kivimäki M. Antidepressant medication use and future risk of cardiovascular disease: the Scottish Health Survey. Eur Heart J. 2011;32(4):437–42.

158. Scherrer JF, Garfield LD, Lustman PJ, Hauptman PJ, Chrusciel T, Zeringue A, et al. Antidepressant drug compliance: reduced risk of MI and mortality in depressed patients. Am J Med. 2011;124(4):318–24.

159. Boehm JK, Peterson C, Kivimäki M, Kubzansky L. A prospective study of positive psychological well-being and coronary heart disease. Health Psychol. 2011;30(3):259–67.

160. Kubzansky LD, Thurston RC. Emotional vitality and incident coronary heart disease: benefits of healthy psychological functioning Arch Gen Psychiatry. 2007;64(12):1393–401.

161. Davidson KW, Mostofsky E, Whang W. Don't worry, be happy: positive affect and reduced 10-year incident coronary heart disease: The Canadian Nova Scotia Health Survey. Eur Heart J. 2010.

162. Consedine NS, Magai C, Bonanno GA. Moderators of the emotion inhibition-health relationship: A review and research agenda. Rev Gen Psychol. 2002;6(2):204–28.

163. de Ridder D, Geenen R, Kuijer R, van Middendorp H. Psychological adjustment to chronic disease. Lancet. 2008;372(9634):246–55.

164. Haynes SG, Feinleib M, Kannel WB. The relationship of psychosocial factors to coronary heart disease in the Framingham study: III. Eight-year incidence of coronary heart disease. Am J Epidemiol. 1980;111(1):37–58.

165. Kubzansky LD, Park N, Peterson C, Vokonas P, Sparrow D. Healthy psychological functioning and incident coronary heart disease: the importance of self-regulation. Arch Gen Psychiatry. 2011;68(4):400–8.

166. Haukkala A, Konttinen H, Laatikainen T, Kawachi I, Uutela A. Hostility, anger control, and anger expression as predictors of cardiovascular disease. Psychosom Med. 2010;72(6):556–62.

167. Appleton AA, Loucks E, Buka SL, Kubzansky LD. Divergent associations of antecedent and response focused emotion regulation strategies with midlife cardiovascular disease risk. Ann Behav Med. in press.

168. Rasmussen K, Ravnsbaek J, Funch-Jenson P, Bagger JP. Oesophageal spasm in patients with coronary artery spasm. Lancet. 1986;1(8474):174–6.

169. Tofler GH, Muller JE. Triggering of acute cardiovascular disease and potential preventive strategies. Circulation. 2006;114(17):1863–72.

170. Mittleman MA, Maclure M, Sherwood JB, Mulry RP, Tofler GH, Jacobs SC, et al. Triggering of acute myocardial infarction onset by episodes of anger. Determinants of Myocardial Infarction Onset Study Investigators [see comments]. Circulation. 1995;92(7):1720–5.

171. Wittstein IS, Thiemann DR, Lima JA, Baughman KL, Schulman SP, Gerstenblith G, et al. Neurohumoral features of myocardial stunning due to sudden emotional stress. N Engl J Med. 2005;352(6):539–48.

172. Wittstein IS. The broken heart syndrome. Cleve Clin J Med. 2007;74(Suppl 1):S17–22.

173. Bounhoure JP. Takotsubo or stress cardiomyopathy. Cardiovasc Psychiatry Neurol. 2012;2012:637672.

174. Wittstein IS. Stress cardiomyopathy: a syndrome of catecholamine-mediated myocardial stunning? Cell Mol Neurobiol. 2012;32(5):847–57.

175. Hayward C. Psychiatric illness and cardiovascular disease risk. Epidemiol Rev. 1995;17:129–38.

176. Noyes R, Clancy J, Hoenk PR, Slymen DR. Anxiety neurosis and physical illness. Compr Psychiatry. 1978;19:407–13.

177. Yan LL, Liu K, Matthews KA, Daviglus ML, Ferguson TF, Kiefe CI. Psychosocial factors and risk of hypertension: the Coronary Artery Risk Development in Young Adults (CARDIA) study. JAMA. 2003;290(16):2138–48.

178. Jonas BS, Franks P, Ingram DD. Are symptoms of anxiety and depression risk factors for hypertension? Longitudinal evidence from the National Health and Nutrition Examination Survey I Epidemiologic Follow-Up Study. Arch Fam Med. 1997;6:43–9.

179. Shinn EH, Poston WS, Kimball KT, St Jeor ST, Foreyt JP. Blood pressure and symptoms of depression and anxiety: a prospective study. Am J Hypertens. 2001;14(7 Pt 1):660–4.

180. Hildrum B, Romild U, Holmen J. Anxiety and depression lowers blood pressure: 22-year follow-up of the population based HUNT study, Norway. BMC Public Health. 2011;11:601.

181. Niu K, Hozawa A, Awata S, Guo H, Kuriyama S, Seki T, et al. Home blood pressure is associated with depressive symptoms in an elderly population aged 70 years and over: a population-based, cross-sectional analysis. Hypertens Res. 2008;31(3):409–16.

182. Nabi H, Chastang JF, Lefevre T, Dugravot A, Melchior M, Marmot MG, et al. Trajectories of depressive episodes and hypertension over 24 years: the Whitehall II prospective cohort study. Hypertension. 2011;57(4):710–6.

183. Scuteri A. Depression and cardiovascular risk: does blood pressure play a role? J Hypertens. 2008;26(9):1738–9.

184. Knol MJ, Twisk JW, Beekman AT, Heine RJ, Snoek FJ, Pouwer F. Depression as a risk factor for the onset of type 2 diabetes mellitus: a meta-analysis. Diabetologia. 2006;49(5):837–45.

185. Mezuk B, Eaton WW, Albrecht S, Golden SH. Depression and type 2 diabetes over the lifespan: a meta-analysis. Diabetes Care. 2008;31(12):2383–90.

186. Rotella F, Mannucci E. Depression as a risk factor for diabetes: a meta-analysis of longitudinal studies. J Clin Psychiat. 2013;74(1):31–7.

187. Rotella F, Mannucci E. Diabetes mellitus as a risk factor for depression: a meta-analysis of longitudinal studies. Diabetes Res Clin Pract. 2013;99(2):98–104.

188. Engum A. The role of depression and anxiety in onset of diabetes in a large population-based study. J Psychosom Res. 2007;62(1):31–8.

189. Golden SH, Williams JE, Ford DE, Yeh HC, Sanford CP, Nieto FJ, et al. Anger temperament is modestly associated with the risk of type 2 diabetes mellitus: the Atherosclerosis Risk in Communities Study. Psychoneuroendocrinology. 2006;31(3):325–32.

190. Pouwer F, Kupper N, Adriaanse MC. Does emotional stress cause type 2 diabetes mellitus? A review from the European Depression in Diabetes (EDID) Research Consortium. Discov Med. 2010;9(45):112–8.

191. Goldbacher EM, Bromberger J, Matthews KA. Lifetime history of major depression predicts the development of the metabolic syndrome in middle-aged women. Psychosom Med. 2009;71(3):266–72.

192. Luppino FS, de Wit LM, Bouvy PF, Stijnen T, Cuijpers P, Penninx BW, et al. Overweight, obesity, and depression: a systematic review and meta-analysis of longitudinal studies. Arch Gen Psychiatry. 2010;67(3):220–9.

193. Pan A, Keum N, Okereke OI, Sun Q, Kivimäki M, Rubin RR, et al. Bidirectional association between depression and metabolic syndrome: a systematic review and meta-analysis of epidemiological studies. Diabetes Care. 2012;35(5):1171–80.

194. Goldbacher EM, Matthews KA. Are psychological characteristics related to risk of the metabolic syndrome? A review of the literature. Ann Behav Med. 2007;34(3):240–52.

195. Hawkins MA, Stewart JC. Do negative emotional factors have independent associations with excess adiposity? J Psychosom Res. 2012;73(4):243–50.

196. Perkonigg A, Owashi T, Stein MB, Kirschbaum C, Wittchen HU. Posttraumatic stress disorder and obesity: evidence for a risk association. Am J Prev Med. 2009;36(1):1–8.

197. Kubzansky LD, Bordelois P, Jun HJ, Roberts AL, Cerda M, Bluestone N, et al. The weight of traumatic stress: a prospective study of posttraumatic stress disorder symptoms and weight status in women. JAMA Psychiatry. 2014;71(1):44–51.

198. Goldberg RJ. Coronary heart disease: epidemiology and risk factors. In: Ockene IS, Ockene JK, editors. Prevention of coronary heart disease. Boston: Little, Brown; 1992. pp. 3–41.

199. Matthews KA, Owens JF, Kuller LH, Sutton-Tyrrell K, Jansen-McWilliams L. Are hostility and anxiety associated with carotid atherosclerosis in healthy postmenopausal women? Psychosom Med. 1998;60(5):633–8.

200. Greenland P, Bonow RO, Brundage BH, Budoff MJ, Eisenberg MJ, Grundy SM, et al. ACCF/AHA 2007 clinical expert consensus document on coronary artery calcium scoring by computed tomography in global cardiovascular risk assessment and in evaluation of patients with chest pain: a report of the American College of Cardiology Foundation Clinical Expert Consensus Task Force (ACCF/AHA Writing Committee to Update the 2000 Expert Consensus Document on Electron Beam Computed Tomography) developed in collaboration with the Society of Atherosclerosis Imaging and Prevention and the Society of Cardiovascular Computed Tomography. J Am Coll Cardiol. 2007;49(3):378–402.

201. Seldenrijk A, van Hout HP, van Marwijk HW, de Groot E, Gort J, Rustemeijer C, et al. Carotid atherosclerosis in depression and anxiety: associations for age of depression onset. World J Biol Psychiatry. 2011;12(7):549–58.

202. Stewart JC, Janicki DL, Muldoon MF, Sutton-Tyrrell K, Kamarck TW. Negative emotions and 3-year progression of subclinical atherosclerosis. Arch Gen Psychiatry. 2007;64(2):225–33.

203. Elovainio M, Keltikangas-Jarvinen L, Kivimäki M, Pulkki L, Puttonen S, Heponiemi T, et al. Depressive symptoms and carotid artery intima-media thickness in young adults: the Cardiovascular Risk in Young Finns Study. Psychosom Med. 2005;67(4):561–7.

204. Ohira T, Diez Roux AV, Polak JF, Homma S, Iso H, Wasserman BA. Associations of anger, anxiety, and depressive symptoms with carotid arterial wall thickness: the multi-ethnic study of atherosclerosis. Psychosom Med. 2012;74(5):517–25.

205. Rozanski A, Gransar H, Kubzansky LD, Wong N, Shaw L, Miranda-Peats R, et al. Do psychological risk factors predict the presence of coronary atherosclerosis? Psychosom Med. 2011;73(1):7–15.

206. Paterniti S, Zureik M, Ducimetiere PJT, Feve JM, Alperovitch A. Sustained anxiety and 4-year progression of carotid atherosclerosis. Arteriosclerosis Thrombosis and Vascular Biology. 2001;21:136–41.

207. Raikkonen K, Matthews KA, Sutton-Tyrrell K, Kuller LH. Trait anger and the metabolic syndrome predict progression of carotid atherosclerosis in healthy middle-aged women. Psychosom Med. 2004;66(6):903–8.

208. Temoshok L. Personality, coping style, emotion, and cancer: toward an integrative model. Cancer Surv. 1987;6:545–67.

209. Grossarth-Maticek R, Kanazir DT, Schmit P, Vetter H. Psychosocial and organic variables as predictors of lung cancer, cardiac infarct and apoplexy: some differential predictors. Pers Indiv Differ. 1985;6:313–21.

210. Grossarth-Maticek R, Eyesenck HJ. Personality, stress and disease. Psychol Rep. 1990;66:355–73.

211. Lemogne C, Consoli SM, Geoffroy-Perez B, Coeuret-Pellicer M, Nabi H, Melchior M, et al. Personality and the risk of cancer: a 16-year follow-up study of the GAZEL cohort. Psychosom Med. 2013;75(3):262–71.

212. Ranchor AV, Sanderman R, Coyne JC. Invited commentary: personality as a causal factor in cancer risk and mortality—time to retire a hypothesis? Am J Epidemiol. 2010;172(4):386–8.

213. Fox BH, Temoshok L, Dreher H. Mind-body and behavior in cancer incidence. Advances. 1988;5(4):41–56.

214. Kiecolt-Glaser JK, Glaser R. Psychoneuroimmunology and cancer: fact or fiction? Eur J Cancer. 1999;35(11):1603–7.

215. Thaker PH, Han LY, Kamat AA, Arevalo JM, Takahashi R, Lu C, et al. Chronic stress promotes tumor growth and angiogenesis in a mouse model of ovarian carcinoma. Nat Med. 2006;12(8):939–44.

216. Feng Z, Liu L, Zhang C, Zheng T, Wang J, Lin M, et al. Chronic restraint stress attenuates p53 function and promotes tumorigenesis. Proc Natl Acad Sci U S A. 2012;109(18):7013–8.

217. Pinquart M, Duberstein PR. Depression and cancer mortality: a meta-analysis. Psychol Med. 2010;40(11):1797–810.

218. Mitchell AJ, Chan M, Bhatti H, Halton M, Grassi L, Johansen C, et al. Prevalence of depression, anxiety, and adjustment disorder in oncological, haematological, and palliative-care settings: a meta-analysis of 94 interview-based studies. Lancet Oncol. 2011;12(2):160–74.

219. Kroenke CH, Bennett GG, Fuchs C, Giovannucci E, Kawachi I, Schernhammer E, et al. Depressive symptoms and prospective incidence of colorectal cancer in women. Am J Epidemiol. 2005;162(9):839–48.

220. Chida Y, Hamer M, Wardle J, Steptoe A. Do stress-related psychosocial factors contribute to cancer incidence and survival? Nat Clin Pract Oncol. 2008;5(8):466–75.

221. Anderson BL, Kiecolt-Glaser JK, Glaser R. A biobehavioral model of cancer stress and disease course. Am Psychol. 1994;49(5):389–404.

222. Possel P, Adams E, Valentine JC. Depression as a risk factor for breast cancer: investigating methodological limitations in the literature. Cancer Cause Control. 2012;23(8):1223–9.

223. Eskelinen M, Ollonen P. Beck Depression Inventory (BDI) in patients with breast disease and breast cancer: a prospective case-control study. In Vivo. 2011;25(1):111–6.

224. Chida Y, Steptoe A. Positive psychological well-being and mortality: a quantitative review of prospective observational studies. Psychosom Med. 2008;70(7):741–56.

225. Aspinwall LG, Tedeschi RG. The value of positive psychology for health psychology: progress and pitfalls in examining the relation of positive phenomena to health. Ann Behav Med. 2010;39(1):4–15.

226. Coyne JC, Tennen H. Positive psychology in cancer care: bad science, exaggerated claims, and unproven medicine. Ann Behav Med. 2010;39(1):16–26.

227. Cohen S, Doyle WJ, Skoner DP, Fireman P, Gwaltney JM, Newsom JT. State and trait negative affect as predictors of objective and subjective symptoms of respiratory viral infections. J Pers Soc Psychol. 1995;68(1):159–69.

228. Kiecolt-Glaser JK, Glaser R. Psychological influences on immunity: implications for AIDS. Am Psychol. 1988;43:892–8.

229. O'Leary A. Stress, emotion, and human immune function. Psychol Bull. 1990;108(363–382).

230. Nakata A, Irie M, Takahashi M. Psychological distress, depressive symptoms, and cellular immunity among healthy individuals: a 1-year prospective study. Int J Psychophysiol. 2011;81(3):191–7.

231. Cohen S, Doyle WJ, Turner RB, Alper CM, Skoner DP. Emotional style and susceptibility to the common cold. Psychosom Med. 2003;65:652–7.

232. Cohen S, Tyrrell DAJ, Smith AP. Negative life events, perceived stress, negative affect, and susceptibility to the common cold. J Pers Soc Psychol. 1993;64:131–40.

233. Cohen S, Alper CM, Doyle WJ, Treanor JJ, Turner RB. Positive emotional style predicts resistance to illness after experimental exposure to rhinovirus or influenza a virus. Psychosom Med. 2006;68(6):809–15.

234. Segerstrom SC, Hardy JK, Evans DR, Greenberg RN. Vulnerability, distress, and immune response to vaccination in older adults. Brain Behav Immun. 2012;26(5):747–53.

235. Glaser R, Robles TF, Sheridan J, Malarkey WB, Kiecolt-Glaser JK. Mild depressive symptoms are associated with amplified and prolonged inflammatory responses after influenza virus vaccination in older adults. Arch Gen Psychiatry. 2003;60(10):1009–14.

236. Von Korff M, Alonso J, Ormel J, Angermeyer M, Bruffaerts R, Fleiz C, et al. Childhood psychosocial stressors and adult onset arthritis: broad spectrum risk factors and allostatic load. Pain. 2009; 143(1–2):76–83.

237. Scott KM, Von Korff M, Angermeyer MC, Benjet C, Bruffaerts R, de Girolamo G, et al. Association of childhood adversities and early-onset mental disorders with adult-onset chronic physical conditions. Arch Gen Psychiatry. 2011;68(8):838–44.

238. Goodwin RD, Sourander A, Duarte CS, Niemela S, Multimaki P, Nikolakaros G, et al. Do mental health problems in childhood predict chronic physical conditions among males in early adulthood? Evidence from a community-based prospective study. Psychol Med. 2009;39(2):301–11.

239. Chida Y, Vedhara K. Adverse psychosocial factors predict poorer prognosis in HIV disease: a meta-analytic review of prospective investigations. Brain Behav Immun. 2009;23(4):434–45.

240. Chida Y, Mao X. Does psychosocial stress predict symptomatic herpes simplex virus recurrence? A meta-analytic investigation on prospective studies. Brain Behav Immun. 2009;23(7):917–25.

241. Wang G, Zhou T, Wang L, Fu JJ, Zhang HP, Ji YL. Relationship between current psychological symptoms and future risk of asthma outcomes: a 12-month prospective cohort study. J Asthma. 2011;48(10): 1041–50.

242. Kanaan RA, Lepine JP, Wessely SC. The association or otherwise of the functional somatic syndromes. Psychosom Med. 2007;69(9):855–9.

243. Sharpe M, Carson A. "Unexplained" somatic symptoms, functional syndromes, and somatization: do we need a paradigm shift? Ann Intern Med. 2001;134(9 Pt 2):926–30.

244. Yunus MB. Fibromyalgia and overlapping disorders: the unifying concept of central sensitivity syndromes. Semin Arthritis Rheum. 2007;36(6):339–56.

245. Maunder RG, Levenstein S. The role of stress in the development and clinical course of inflammatory bowel disease: epidemiological evidence. Curr Mol Med. 2008;8(4):247–52.

246. Glombiewski JA, Sawyer AT, Gutermann J, Koenig K, Rief W, Hofmann SG. Psychological treatments for fibromyalgia: a meta-analysis. Pain. 2010;151(2):280–95.

247. Gurmankin Levy A, Kubzansky LD, Maselko J, Richman L, Bauer M. Why do those with an anxiety disorder utilize more non-mental health care than those without? Health Psychol. 2007;26:333–40.

248. Walker J, Sharpe M, Wessely S. Commentary: symptoms not associated with disease: an unmet public health challenge. Int J Epidemiol. 2006;35(2):477–8.

249. Angell M. Disease as a reflection of the psyche. N Engl J Med. 1985;312:1570–2.

250. Relman AS, Angell M. Resolved: Psychosocial interventions can improve clinical outcomes in organic disease (con). Psychosom Med. 2002;64(4):558–63.

251. Sloan EK, Capitanio JP, Tarara RP, Mendoza SP, Mason WA, Cole SW. Social stress enhances sympathetic innervation of primate lymph nodes: mechanisms and implications for viral pathogenesis. J Neurosci. 2007;27(33):8857–65.

252. Cole SW. Social regulation of human gene expression. Curr Dir Psychol Sci. 2009;18(3):132–7.

253. Kessler D. The end of overeating: Taking control of the insatiable American appetite. New York: Rodale Press; 2009.

254. Epel ES, Blackburn EH, Lin J, Dhabhar FS, Adler NE, Morrow JD, et al. Accelerated telomere shortening in response to life stress. Proc Natl Acad Sci U S A. 2004;101(49):17312–5.

255. Aubert G, Lansdorp PM. Telomeres and aging. Physiol Rev. 2008;88(2):557–79.

256. Simon NM, Smoller JW, McNamara KL, Maser RS, Zalta AK, Pollack MH, et al. Telomere shortening and mood disorders: preliminary support for a chronic stress model of accelerated aging. Biol Psychiatry. 2006;60(5):432–5.

257. Depue RA, Morrone-Strupinsky JV. A neurobehavioral model of affiliative bonding: implications for conceptualizing a human trait of affiliation. Behav Brain Sci. 2005;28(3):313–50; discussion 50–95.

258. Insel TR, Young LJ. The neurobiology of attachment. Nat Rev Neurosci. 2001;2(2):129–36.

259. Singer B, Friedman E, Seeman T, Fava GA, Ryff CD. Protective environments and health status: cross-talk between human and animal studies. Neurobiol Aging. 2005;26(Suppl 1):113–8.

260. Heinrichs M, Baumgartner T, Kirschbaum C, Ehlert U. Social support and oxytocin interact to suppress cortisol and subjective responses to psychosocial stress. Biol Psychiatry. 2003;54(12):1389–98.

261. Kirsch P, Esslinger C, Chen Q, Mier D, Lis S, Siddhanti S, et al. Oxytocin modulates neural circuitry for social cognition and fear in humans. J Neurosci. 2005;25(49):11489–93.

262. Hoge EA, Pollack MH, Kaufman RE, Zak PJ, Simon NM. Oxytocin levels in social anxiety disorder. CNS Neurosci Ther. 2008;14(3):165–70.

263. Kubzansky LD, Mendes WB, Appleton AA, Block J, Adler GK. A heartfelt response: oxytocin effects on response to social stress in men and women. Biol Psychol. 2012;90(1):1–9.

264. Miller GE, Chen E, Sze J, Marin T, Arevalo JM, Doll R, et al. A functional genomic fingerprint of chronic stress in humans: blunted glucocorticoid and increased NF-kappaB signaling. Biol Psychiatry. 2008.

265. Dusek JA, Otu HH, Wohlhueter AL, Bhasin M, Zerbini LF, Joseph MG, et al. Genomic counter-stress changes induced by the relaxation response. PLoS One. 2008;3(7):e2576.

266. Cole SW, Hawkley LC, Arevalo JM, Sung CY, Rose RM, Cacioppo JT. Social regulation of gene expression in human leukocytes. Genome Biol. 2007;8(9):R189.

267. Rozanski A, Blumenthal JA, Davidson KW, Saab PG, Kubzansky L. The epidemiology, pathophysiology, and management of psychosocial risk factors in cardiac practice: the emerging field of behavioral cardiology. J Am Coll Cardiol. 2005;45(5):637–51.

268. Fredrickson BL, Mancuso RA, Branigan C, Tugade MM. The undoing effect of positive emotions. Motiv Emotion. 2000;24(4):237–58.

269. Fredrickson BL, Levenson RW. Positive emotions speed recovery from the cardiovascular sequelae of negative emotions. Cogn Emot. 1998;12(2):191–220.

270. Kuh D, Ben-Shlomo Y. A life course approach to chronic disease epidemiology. New York: Oxford University Press; 1997.

271. National Research Council and Institute of Medicine. From neurons to neighborhoods: the science of early child development. Committee on integrating the science of early childhood development. Shonkoff JP, Phillips DA, editors. Washington, DC: National Academy Press; 2000.

272. Lumley MA. Alexithymia, emotional disclosure, and health: a program of research. J Pers. 2004;72(6): 1271–300.

273. Frattaroli J. Experimental disclosure and its moderators: a meta-analysis. Psychol Bull. 2006;132(6):823–65.

274. Meads C, Nouwen A. Does emotional disclosure have any effects? A systematic review of the literature with meta-analyses. Int J Technol Assess Health Care. 2005;21(2):153–64.

275. Caspi A. The child is father of the man: personality continuities from childhood to adulthood. J Pers Soc Psychol. 2000;78(1):158–72.

276. Kubzansky LD, Martin LT, Buka SL. Early manifestations of personality and adult emotional functioning. Emotion. 2004;4(4):364–77.

277. Mischel W, Ayduk O, Berman MG, Casey BJ, Gotlib IH, Jonides J, et al. "Willpower" over the life span: decomposing self-regulation. Soc Cogn Affect Neurosci. 2011;6(2):252–6.

278. Appleton AA, Buka SL, McCormick MC, Koenen KC, Loucks EB, Gilman SE, et al. Emotional functioning at age 7 years is associated with C-reactive protein in middle adulthood. Psychosom Med. 2011;73(4):295–303.

279. Slopen N, Kubzansky LD, Koenen KC. Internalizing and externalizing behaviors predict elevated inflammatory markers in childhood. Psychoneuroendocrinology. 2013;38(12):2854–62.

280. Macleod J, Davey Smith G, Heslop P, Metcalfe C, Carroll D, Hart C. Psychological stress and cardiovascular disease: empirical demonstration of bias in a prospective observational study of Scottish men. Br Med J. 2002;324:1247–54.

281. Glassman AH, Bigger JT, Gaffney M, Shapiro PA, Swenson JR. Onset of major depression associated with acute coronary syndromes: relationship of onset, major depressive disorder history, and episode severity to sertraline benefit. Arch Gen Psychiatry. 2006;63(3):283–8.

282. Craddock N, O'Donovan MC, Owen MJ. Genome-wide association studies in psychiatry: lessons from early studies of non-psychiatric and psychiatric phenotypes. Mol Psychiatry. 2008;13(7):649–53.

CHAPTER 10

CHANGING HEALTH BEHAVIORS IN A SOCIAL CONTEXT

Cassandra Okechukwu, Kirsten Davison, and Karen Emmons

Thinking is easy, acting is difficult, and to put one's thoughts into action is the most difficult thing in the world.

INTRODUCTION

BEHAVIOR MATTERS

Health behaviors play a strong role in shaping population health outcomes for most major diseases in the United States, in other developed countries, and, increasingly, in the developing world. Behavior impacts the leading causes of morbidity and mortality, and its role extends to the development and progression of diseases, the effectiveness of treatments, and quality of life (1). The most common diseases can be prevented with the adoption of healthier behaviors (2–4). Smoking is a chief contributor to morbidity and the leading cause of mortality (3). The second leading cause of mortality, obesity, is driven by diet and physical activity throughout the lifecourse (5, 6). Taken together, smoking, physical inactivity, and diet cause or exacerbate ten of the fifteen leading causes of death in developed nations (3, 7). The importance of health behaviors is underscored by the heavy focus on prevention in the US Affordable Health Care Act, which provides unprecedented universal coverage for preventive health services, including behavioral interventions (8).

The complexities of health behavior, and its interaction with social and physical environmental contexts, are clear and were documented as early as the 1800s. Villermé noted that social factors, such as the standards of living and duration of work, constrained the behaviors of people in different trades, such that different professions were associated with particular behavioral tendencies

and even levels of hygiene (9). Since then, social epidemiology has highlighted the important contribution of social determinants to morbidity and mortality. In this chapter, we use a social determinants lens to examine the contribution of health behavior and health behavior interventions to population health. Given the continued debate about the return on investments for health behavior interventions, we summarize the evolution of health behavior interventions and discuss key innovations in the field of health behavior change. It will be argued that behavioral change interventions have evolved substantially from an exclusive focus on individual-based approaches to those that integrate individual- and population-based strategies, typically using multilevel intervention strategies. It is further argued that without attention to behavior change, there is little likelihood of meaningful improvement in population health. Woven throughout the discussion is the importance of social context to population trends in adopting and maintaining health behaviors, as well as to consequences of unhealthy behaviors. Although an extensive social epidemiological literature addresses the relationship between social factors and health outcomes, little attention has been paid to using social epidemiology to inform health behavior interventions. Therefore, we conclude by advocating for a shift toward solution-oriented social epidemiology that collaborates with behavioral scientists to redress the unequal distribution of health behaviors and health outcomes in the population.

THE SOCIAL PATTERNING OF HEALTH BEHAVIOR

TOBACCO USE

Despite the continued role of tobacco as a leading cause of death and disability, significant progress has been made. The adult smoking prevalence in the United States declined steadily from above 50% in the 1950s to around 19% in 2011 (10, 11); the Healthy People 2020 goal is 12%. The CDC calculated that there has been a 42.4% decline in the prevalence of smoking among those aged ≥18 years since 1965, but the declines slowed between 2004 and 2010 (12). An estimated 43.8 million people currently smoke cigarettes in the United States (10). Among high school students, smoking increased from 27.5% in 1991 to 36.4% in 1997, but has since declined steadily to reach 18.1% in 2011; the Healthy People 2020 prevalence goal for those under 18 is 16% (13).

Starting in the late 1950s, around the time that its impact on health was first documented, smoking became associated with social and economic disadvantage, an association that has persisted for over five decades. In the general US population, the widest disparities in smoking prevalence exist by educational attainment and by occupational groupings. Only 9.3% of adults who have a college degree and 5% of those with a postgraduate degree are smokers. In contrast, the prevalence of smoking among adults with a high school diploma, GED, or some high school is 23.8%, 45.3%, and 34.6%, respectively (10). Although the exact history of the disparities has been difficult to understand due to data limitations, for the past two decades, smoking prevalence among American Indian/Alaskan Native (AI/AN) populations has been at least 10% higher than among whites, the next highest prevalence group by race/ethnicity (e.g., AI/AN 2011 prevalence was 31.5%, while it was 20.6% for whites) (13). The continued high prevalence of smoking and

lower proportion of smoking cessation attempts and successes among AI/AN populations in the United States is a public health inequity that has not received the specialized attention given to other marginalized groups with high prevalence (14–17). A clear gradient in smoking initiation, prevalence, amount smoked, and quit success by educational attainment exists among African American and white populations (18).

The prevalence of smoking among men (21.6%) continues to be higher than the prevalence among women (16.5%) (10). Unemployment, which has an inverse relationship with education, is associated with higher prevalence of smoking for people in the labor force (19, 20). Blue-collar construction and mining workers, most of whom have a high school education or less, continue to have the highest smoking prevalence by occupation (31%) while the lowest prevalence by occupation is among those in education, training, and library occupations (8.7%) (12). Generally, blue-collar and service workers are more likely to be ever smokers, current daily users, and heavier smokers than are white-collar workers (16, 21–26).

PHYSICAL ACTIVITY, DIET, AND OBESITY

The decrease in US smoking prevalence in the last few decades has coincided with an increase in obesity rates that started in the later part of the twentieth century (27, 28). More than one-third of US adults (35.7%) and a significant proportion of children and adolescents (16.9%) are obese (28). The rise in obesity corresponds to increased caloric intake and sedentary behaviors. A varied and healthful diet high in fruits, vegetables, and grains, and sufficient physical activity are essential for prevention of chronic diseases and maintenance of healthy weight. However, less than 30% of the US population consumed five or more servings of fruits and vegetables a day as recommended by the USDA (29, 30). Moreover, adherence to dietary and physical activity recommendations, along with rates of obesity, varies by income, education, and race/ethnicity.

Over the past two decades, obesity has increasingly become a problem with significant patterning by race/ethnicity. At 49.5% and 40%, respectively, obesity rates for blacks and Hispanics are higher than the rate for non-Hispanic whites, which is 34.3% (27). The effects of obesity extend to future generations, with 17% of children and adolescents in the United States classified as obese, and black and Hispanic children impacted more than other groups. Race/ethnicity- and socioeconomic status (SES)–based disparities in pediatric obesity have not changed in the last decade (31).

Consistent income differences in the proportion of adults meeting dietary recommendations have been observed. Overall, adults in the highest income group show greater adherence to dietary recommendations compared with adults in lower income groups (32). For some foods, including total vegetables, milk, and oil consumption, almost twice the percentage of highest-income adults meet recommendations, compared with lowest income adults. Links between family income and dietary patterns are more mixed among children. Children in the highest income group show the highest level of compliance with dietary recommendations for whole fruit and whole grains. However, children in the lowest income group are most likely to meet recommendations for total vegetables, starchy vegetables, meat, and beans (32). Mixed results by income among children may be explained by low-income parents buffering their children from severe dietary compromises in

the context of food insecurity, as well as by access to food assistance programs, which align with the 2005 dietary recommendations.

Dietary disparities are more pronounced by race and ethnicity. According to the Behavioral Risk Factor Surveillance System 2009 and the National Health and Nutrition Examination Study (NHANES 2001–2004), significantly fewer non-Hispanic black adults met dietary recommendations compared with Hispanic (Mexican) and non-Hispanic white adults. In particular, non-Hispanic black adults were least likely to consume recommended amounts of whole fruits, total vegetables, total grains, and milk, and exceeded recommended added sugar intake (32, 33). Similar patterns have been observed among children, such that fewer non-Hispanic black children consumed recommended levels of whole fruits, other vegetables, total grains, and milk compared with non-Hispanic white children (32).

When self-report measures of physical activity are utilized, racial/ethnic and income disparities in the achievement of the physical activity recommendations mirror those observed for dietary recommendations (34–36). A different pattern is observed, however, when accelerometers, which provide an objective measure of physical activity, are utilized. Based on accelerometry data from NHANES 2003–2004, non-Hispanic black children aged 6 to 11 years showed greatest adherence to the physical activity recommendations (PA) compared with non-Hispanic white or Mexican American children (35). Few consistent differences in meeting PA recommendations among children were observed by family income. Among adults, greatest compliance with physical activity recommendations was observed among Mexican Americans and individuals with less than a high school education, compared with their respective counterparts (34). Surprisingly high rates of physical activity among non-Hispanic black children and Mexican American and low-educated adults may be explained by higher rates of transportation-related physical activity (e.g., walking) and occupation-related physical activity, respectively (34, 35). One consistent pattern across self-report and objective measures of physical activity is that men achieve higher adherence to PA recommendations than women.

SOCIAL CONTEXT AND HEALTH BEHAVIOR

The differential distribution of health behaviors by population social and economic characteristics oftentimes reflects variation in the social contexts in which individuals live their daily lives. Historical and contemporary social structures shape people's day-to-day experiences in ways that are typically not seen when examining health behaviors by broad categorizations of race/ethnicity, sexual orientation, income, and occupation. The physical and social environments surrounding individuals are not randomly distributed. Knowledge of such structures and their interrelationships can help pinpoint intervention targets and more broadly guide the development of efficacious interventions.

Historical events, including discriminatory policies of the US federal government and corporations, shaped the present overrepresentation of African American communities in poor, urban neighborhoods. The population density per food market is much greater in poor neighborhoods compared to middle- and upper-class neighborhoods; the typical cost of food is approximately 15% to 20% higher in poor neighborhoods, while the quality of food available is poorer (37–39).

Several studies have documented that there is less direct marketing of unhealthy products such as alcohol and tobacco in wealthier neighborhoods in the United States (40–42). In contrast, low-income neighborhoods and those with high concentrations of African Americans have more fast food outlets (43–45). Later in this chapter, we discuss how some health behavior interventions have used the information about unequal access to healthy foods and other resources in the design of their interventions.

The accumulation of social disadvantage increases one's risk of many deleterious health behaviors, and reduces the likelihood of successful behavior change. For example, the risk of smoking by a white man without a college degree is 17% higher, and the probability of cessation 10% lower, if he is also in a lower occupational class and is poor (16). Generally, smoking cessation attempts occur at about the same rate for smokers of different race/ethnicity, education, and occupational classes, but disparities begin to emerge in successful smoking cessation, with white, more highly educated, and white-collar smokers having better success (16). This pattern may be at least partially explained by differences in the social context of smoking. For example, blue-collar and service workplaces are less likely to be covered by clean indoor air policies that restrict or ban smoking in workplaces and by health insurance policies that provide coverage for smoking cessation (46–51). When a nationally representative study of workplaces accounted for differences in worksite smoking policies and programs, differences in cessation rates between white- and blue-collar workers disappeared (48). Moreover, the social norms at blue-collar workplaces, and within their family households, provide less pressure to quit smoking and lower support for smoking cessation, factors associated with less motivation to quit (52–57).

The impact of social context on health behaviors is further illustrated in the work of Hillary Graham (58, 59). Focusing on populations in England and Europe, Graham argues that different dynamics drive the smoking habits of low-income women compared with middle- and upper-income women. She identifies four categories of influence: (1) everyday responsibilities (e.g., childcare, caring for other family members) and patterns of paid work; (2) material circumstances (e.g., housing situation, partner's employment, income and benefit status, access to transportation and telephone); (3) social support and social service networks (e.g., relationships with partners, family, and friendship networks; feelings of belonging); and (4) personal and health resources (e.g., physical and psychosocial health, health beliefs, health behaviors, alternative coping strategies). Following an extensive qualitative study and analysis, she concludes that low-income women use smoking as a means of coping with their economic pressures and the resulting demands placed on them to care for others.

In the socioeconomic environments where smoking is concentrated, this behavior can be very adaptive in terms of helping individuals meet the immediate demands of their life circumstances. Smoking is a relatively cheap and easily available source of stress relief, which might be especially salient for those in disadvantaged socioeconomic positions (60, 61). Cessation may be perceived as futile in the face of other health hazards posed by the work and neighborhood environments (62–66). A comparison of the physiological and psychological effects of smoking to the stressors of poverty is presented in Table 10.1 to illustrate how populations who face high stress might view smoking cessation. From a short-term cost-benefit perspective, smoking may in fact be an adaptive behavior in impoverished circumstances, which further highlights the importance of addressing the social context in intervention design and delivery.

TABLE 10.1: The social context of smoking among low-income population

"Effects" of Smoking	Characteristics of Social Environment
Reduces stress	High stress
Relatively low cost	Few economic resources
Provides social connection	Social norms support smoking
Causes disease/death in the long run	Causes disease/death in the short and long run

HEALTH BEHAVIOR CHANGE MODELS THAT CONSIDER THE SOCIAL CONTEXT

Sorensen, Emmons, and colleagues have developed a guiding framework that explicates the role of the social context in tobacco cessation as well as other health behavior change (see Figure 10.1) (67, 68). This framework defines a set of factors and conditions that sit along the pathway of behavior change and cut across multiple levels of influence. Modifying conditions independently impact on outcomes, but are not influenced by the intervention. At the same time, mediating mechanisms are variables in the pathway between the event or intervention and the outcomes. The model draws on mediating mechanisms from a number of social and behavioral theories that are potentially modifiable within the context of intervention in a range of channels. Social context, including life experiences, social relationships, organizational structures, physical environment, and societal influences, may function as either modifying conditions or mediating mechanisms, depending on the location within or outside the causal pathway between the intervention and the outcomes. As described above, social class, race and ethnicity, gender, age, and language are important parameters for examining the distributions of health outcomes and behavioral risk factors. These sociodemographic characteristics shape social context and day-to-day realities, which in turn influence health behaviors and people's ability to enact health behavior change.

A similar argument for the role of social contextual influence can be made for the influence of the social context on children. For example, research suggests that low-income parents may adopt health-compromising approaches to their children's diet and physical activity, and permissive approaches to children's screen use, as a means of coping with stressful family and social circumstances (69). Social contextual factors and their impact on children's obesity risk are articulated in greater detail in the family ecological model (FEM; see Figure 10.2) (69). Examples of family ecological factors can include general family functioning, family cohesion, parents' work demands and schedules, the health needs of family members, access to social support systems inside and outside the family, housing stability, and food security. The general argument presented in the FEM is that family ecological factors shape family social and emotional environments and parents' experiences of social disparities and chronic stress, both of which impact parents' food and physical activity parenting practices, and in turn children's diet and physical activity behaviors and, subsequently, their obesity risk. The implication is that successful family interventions for obesity prevention and control will require a holistic approach that addresses family ecological factors and parents' experiences of social disparities and chronic stress in conjunction with more traditional features of obesity interventions, including knowledge and skill development in healthy lifestyles.

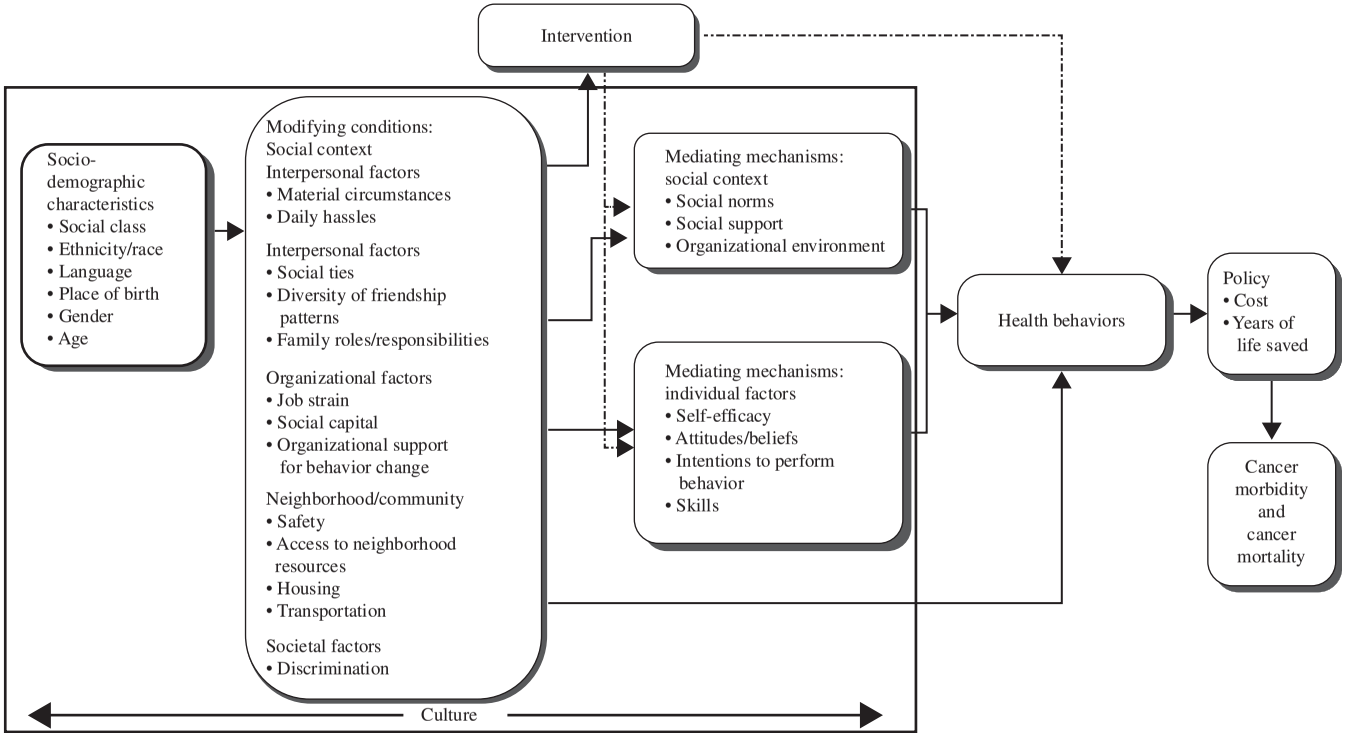


FIGURE 10.1: Social Contextual Framework.

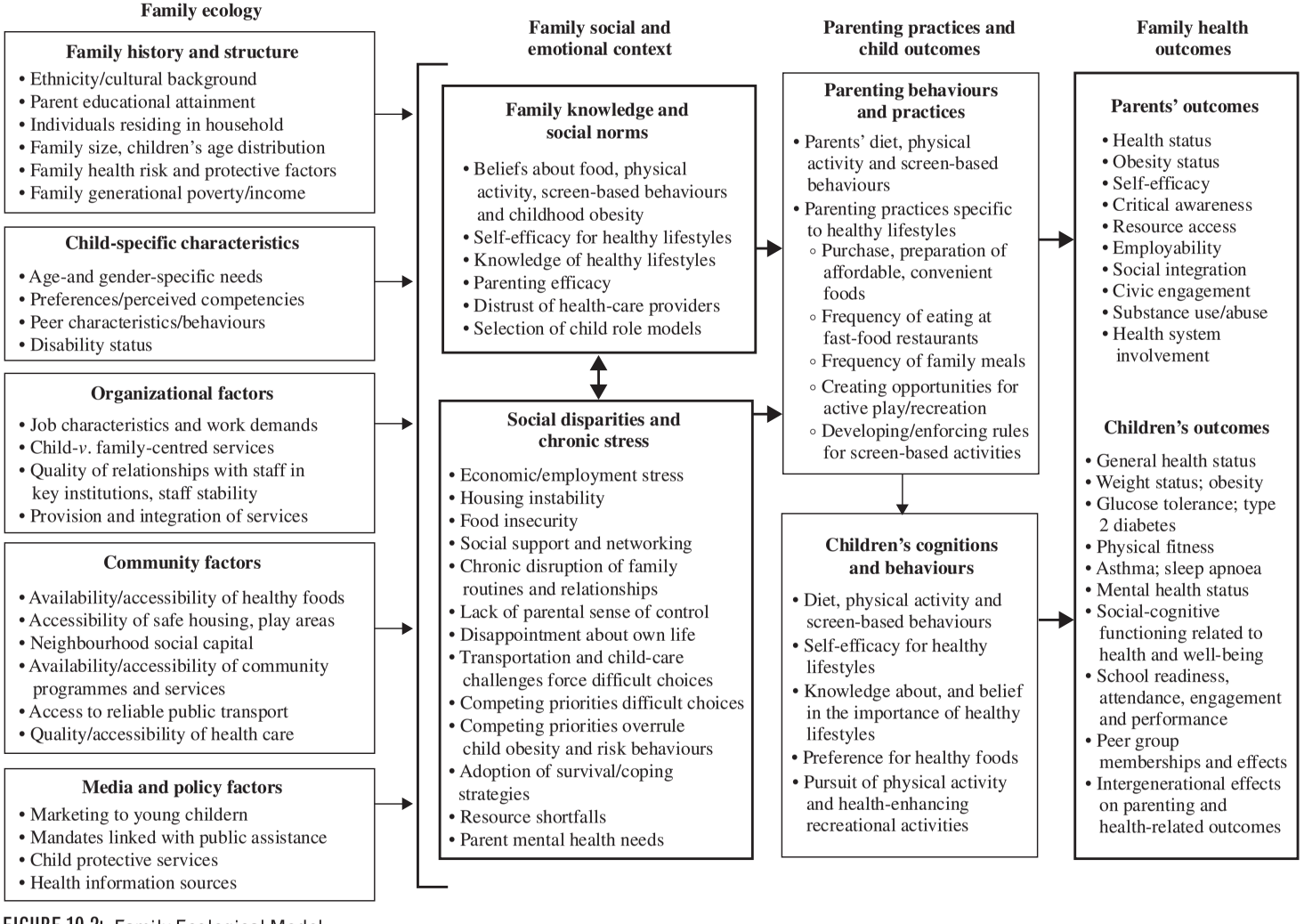


FIGURE 10.2: Family Ecological Model.

As illustrated in this section, behavior change efforts with diverse populations in general, and particularly with those facing substantial health disparities, have to compete with the more immediate financial hardship, caretaking responsibilities, and already existing physical and mental health problems that come with social disadvantage. We now turn our attention to the health behavior change literature, its evolution over time, and the opportunities to address social contextual factors in behavioral interventions.

HEALTH BEHAVIOR CHANGE INTERVENTIONS

Given the extensive impact that health behaviors have on preventable morbidity and mortality, it is critical that we develop evidence-based strategies for reducing these risk factors. This section reviews the evolution of intervention strategies, which has reflected growing knowledge of the importance of social context.

INDIVIDUALLY ORIENTED INTERVENTIONS

Early Approaches

Over the past two to three decades, there has been a substantial evolution in health behavior change interventions. During the 1980s and 1990s, most interventions were individually targeted (70, 71). These interventions employed multiple strategies but typically relied heavily on health education and advice giving and were grounded in theories emphasizing the need to strengthen psychosocial precursors for behavior change, including knowledge and attitudes such as self-efficacy and outcome expectations (70, 72–74). Interventions were typically small-scale, long in duration, and complex. Moreover, they were generally developed in clinical or laboratory settings, with change expected to occur in a linear and rational sequence (75, 76). As a result, many of them were costly, and the external validity of some of the successful individually targeted interventions has been questioned (75–78). An important criticism is that they often targeted highly motivated individuals who were eager for change and who might not have been representative of the general public (77, 79). For the interventions to have population-level impact, they would have to work for both those who are highly motivated and those who are not. Other concerns include the limited reach of intensive individually focused interventions, and difficulty sustaining and scaling such interventions. Furthermore, much of the early behavior change literature had a limited focus on and success with the most vulnerable population groups, and did not address the social and economic circumstances in which health behaviors take place (or in which the individual is embedded).

Motivational Interviewing

One of the significant developments in individually targeted interventions—motivational interviewing—became increasingly popular in the 2000s (80, 81). It involves the use of counselors in nonjudgmental and nonconfrontational encounters to help individuals work through their ambivalence to behavior change (72, 82, 83). The approach emphasizes working through ambivalence by dissecting the positive and negative aspects of behavior change and mapping out a new course of action to

lead to behavior change (83). In this process, aspects of the social context that influence behavior can be addressed. Motivational interviewing has been adapted to several settings, including healthcare settings, and the modalities used have been expanded to include delivery by phone and Internet, and through mailed print materials, videotapes, and computer-generated expert systems (72). Generally, meta-analytic reviews have found that motivational interviewing can be effective for various health behavior outcomes including smoking cessation, regular physical activity, and improved diet, and with a range of population groups (84–87). However, the effectiveness of motivational interviewing is highly dependent on the training of counselors; reviews have found that general practitioners and highly trained counselors are needed for the strategy to be successful (87).

Behavioral Economics: A Recent Extension of Individual Approaches

There has been considerable excitement about the potential of behavioral economics to speed behavior change. Behavioral economics has attracted attention at least in part because of its conceptual appeal and its potential to offer low-cost, unobtrusive solutions—although it still largely targets individual health behaviors—reaching beyond regulatory interventions (e.g., targeted taxes and subsidies), and integrating strategies to help individuals overcome common decision-making biases that lead to poorer health behaviors (e.g., "present bias," or focusing on costs and benefits that are immediate, and underattending to those that are delayed). Behavioral economic interventions are premised on the notion that the same errors that trip people up can also be used to help them; for example, present bias can be overcome by offering small, frequent, and immediate payments for beneficial behaviors (88). Although there are a number of ways in which behavioral economics can be applied to influence health behaviors (see Chapter 13), one strategy that has attracted particular attention in the behavior change community is the use of incentives to promote desired behavior.

Although there have been some positive results from behavioral economic approaches in general and from the use of incentives in particular, the evidence base, particularly with regard to addressing social determinants in health behaviors, is limited (89). Blumenthal and colleagues note that one clear finding from the literature on incentives is that they are more effective at achieving behavior change for preventive care requiring a single activity, such as receiving a vaccination for influenza, than for actions that require ongoing engagement, such as smoking cessation (89), a finding that parallels that found in the behavioral intervention literature (74). The literature targeting population health and disparities is limited, but the evaluation of three state-level Medicaid incentive programs found mixed results (89). For example, in Florida between 2006 and 2011, only about half of the $41.3 million in available credits that were "earned" for a variety of behaviors, from attending wellness visits to smoking cessation, was "claimed" by enrollees. Nearly all credits that were distributed were based on enrollees' attending either a childhood preventive care visit or an adult or child office visit. However, this incentive program was unsuccessful in engaging participants in health education or counseling to reduce risk for chronic disease; over the five years of the program, only two enrollees earned credits for participation in a smoking cessation program, and only two enrollees earned credits for participation in an exercise program. This evaluation assessed only the use of financial incentives, and not other behavioral economic principles. There were also several methodological issues that should be addressed in the design of future programs to assess the impact of behavioral economic strategies on disparities (e.g., incentive complexity, size).

STRATEGIES FOR TAKING BEHAVIORAL INTERVENTIONS TO SCALE

As attention to the importance of a population perspective has developed, there has been an increased focus on interventions that have impact—that is, that combine the best of individually oriented behavioral interventions in a format that can be delivered to broader audiences. Two such approaches that facilitate scale-up include various forms of telephone-based interventions and eHealth interventions.

Telephone-Based Interventions

Early and ongoing efforts focused on telephone-based interventions. A recent structured review of 25 studies (with 27 comparisons) of physical activity and/or diet intervention studies found that 20 of the 27 comparisons provided evidence for initiation of behavior change. One-third of the studies evaluated postintervention maintenance of change; among those, maintenance was achieved for at least half of the outcomes (90). A continued concern, however, is how to take telephone-based interventions to scale due to the training and ongoing labor costs (91). As a result, attention has turned to automated and interactive telephone systems (AVR/IVR). Studies using AVR/IVR systems in combination with case manager follow-up for diabetes among low-income patients have found significant improvements in self-care and glycemic control and high levels of satisfaction (92, 93).

Another way to scale up phone interventions, particularly in nonclinical settings, is through the use of call services, such as smoking cessation quitlines (94–96). Publicly supported quitlines currently exist in North America, Europe, Australia, and New Zealand (94). Reviews have consistently found that quitlines effectively increase short-term and prolonged smoking cessation for different types of callers (pooled odds from 1.41 [95% CI: 1.27, 1.57] to 1.9 [95% CI: 1.7, 2.2]) (96, 97). Quitline protocols are usually based on motivational interviewing, with a pro-active approach in which the smokers make the first call and counselors schedule subsequent calls to first motivate change and then prevent relapse (94, 95). Many US states have adopted quitlines as part of comprehensive tobacco control programs that include population-based approaches to address the social context, such as smoking bans and insurance coverage for smoking cessation programs (94, 98). Services such as quitlines may get a boost from the US federal government through the Affordable Care Act, which provides enhanced federal funding for state Medicaid programs that eliminate cost-sharing for preventive services (8). Also, the Community Transformation Grants that are part of the act provide funds to build partnerships to promote nutrition, physical activity, and smoking cessation (8).

eHealth Interventions

There has also been considerable attention to interactive, or eHealth, technologies, that deliver interventions through one or a combination of mobile websites, text messages, or e-mails. A systematic review of 47 studies of eHealth interventions targeting physical activity and/or dietary behaviors and published between 2000 and 2005 found small effect sizes and limited effectiveness (99). However, the review concluded that eHealth interventions have a great potential for

wider delivery of interventions because of advancements in technology and increased access to technological devices, especially among populations in lower socioeconomic positions and in developing countries. A review of 12 behavior change interventions that utilized only text messages—which are relatively cheap, easy to use, and accessible even to populations in poor countries—found evidence that text messages can be effectively used as a tool for behavior change (100). A promising development is that, although there are still racial and ethnic disparities in access to broadband Internet at home, no such disparities exist in access to mobile phones and smartphones (101, 102). These devices can be harnessed for wider delivery of interventions to populations, such that intervention strategies or modes of intervention that previously targeted the individual can be scaled up to dramatically increase their reach and in turn their public health impact (103). The potential for mobile devices to be used in the delivery of health and healthcare in developing countries is increasingly being recognized even as the science of making them more effective is still evolving (104).

SUMMARY: INDIVIDUAL INTERVENTIONS

Health promotion interventions have increasingly evolved in response to a recognition of their limitations. Individual-level approaches are limited in their potential for health behavior change if they are conducted in isolation without the benefit of interventions and policies that also address societal factors that influence health behaviors. However, such approaches have contributed enormously to our understanding of health behavior, and they play an important role in a comprehensive approach to public health. Lessons learned from individually targeted interventions can be applied to designing effective interventions that can be disseminated more widely. Additionally, as will be discussed in the next section, lessons from the success of some strategies used in the individual approach could be applied to the design of structural interventions that address social contextual issues such as access. Focusing on impact and reach is a more useful dialogue than the increasingly common arguments pitting individual and population approaches against one another. Such a focus would argue for interventions that bridge intervention levels, thus taking advantage of the greater magnitude of individual-level change found in individual interventions, while simultaneously expanding reach into all populations.

COMMUNITY-BASED INTERVENTIONS

Increased recognition that prevention requires efforts beyond the individual level resulted in the development of community-based health promotion interventions (105, 106). Community-based health interventions that utilize the organization as the unit of analysis emerged in the context of large community trials, in workplaces, in schools, and in health care settings. Given the burgeoning understanding of social determinants of health, many community-based interventions targeted social and environmental influences. These interventions were furthered by the expansion of health behavior change theories and frameworks to include community constructs and the emergence and/or popularization of theories that emphasize the social context, such as social cognitive theory (70) and the social contextual model (67).

An important basis of community-based studies, particularly those using principles of community-based participatory research (CBPR), is community participation and ownership in health promotion activities (107, 108). Community-based participatory research engages community members or representatives of the target population actively and equitably in decision-making throughout the research process (109). In doing so, CBPR-informed interventions move away from delivering an intervention "to" members of a target population, to instead "working with" them to address a health problem (110). It is an effective strategy for gaining local knowledge of sociocultural contexts for the development of culturally tailored interventions. Resulting interventions are more likely to leverage individual, institutional, and community assets and provide salient knowledge, skills, and resources to the target population than traditional deficit-focused models (108, 109). This in turn fosters the sustainability of a program.

Community-based population-level approaches have much greater potential for impacting behavior among a larger number of people, although these interventions are typically much less intensive than individually targeted interventions, and therefore the interventions' effects for the individual tend to be much smaller (106). Campbell et al. (2007) have articulated five essential elements of the community-based intervention process that are critical to success (111) and are also critical to addressing social context:

1. careful attention to partnership development and building trust;
2. an everything-on-the-table approach to involving the institution (church, workplace) or community group in recruitment of participants;
3. efforts to understand the cultural/social context through extensive formative research and involvement of key informants or advisors;
4. intervention strategies that incorporate the sociocultural environment and that can be delivered at least in part by the community; and
5. ongoing plans for ensuring program sustainability (leaving something behind).

Studies within community settings have utilized a range of intervention modalities and units of analysis (e.g., individual, family, organization, community). Channel-based interventions have the advantage of having defined populations that can be reached through direct contact, and the opportunity to address physical context through environmental changes. A number of intervention channels, particularly schools and worksites, also provide opportunities to modify the organizational environment and policies as part of intervention. Moreover, community-based interventions increasingly connect interventions across multiple settings, strengthening intervention reach and dose. Examples of settings that have been studied include workplaces, schools, faith-based organizations, physician offices, human services agencies, and housing developments. We have selected the workplace, faith-based, and family settings as examples of community-based interventions, as they illustrate key aspects of these strengths.

WORKSITE INTERVENTIONS

The worksite has emerged as an important medium for delivering contextually relevant behavior change interventions to a wide group of people. As such, worksites are a particularly promising

venue for reaching vulnerable populations, and especially for reaching working-class populations who may not have supportive structures for healthier behaviors at home and in their neighborhoods. A wide range of risk factors have been targeted through the workplace, including smoking (112–115), nutrition (116–118), physical activity (119–121), work-family stress (122, 123), addiction (124–126), cancer screening (127–129), and occupational exposures (130, 131).

In the early days of worksite interventions, workers' health behaviors were treated as risk factors unrelated to the organization of work that, in combination with occupational hazards, led to poor health (132). However, several studies have shown that when workers are trying to enact healthier behaviors, the workplace can play a critical role in supporting or obstructing their progress. Sorensen and colleagues found that workers exposed to occupational hazards were more likely to be smokers (133). In addition, a longitudinal study of 3,606 smokers found that, controlling for social class, probability of smoking cessation decreased with exposure to occupational hazards (134). Sorensen and colleagues integrated the lessons learned about the link between occupational exposure and smoking into their smoking cessation intervention by testing an intervention that integrated efforts to reduce exposure to occupational hazards with interventions to reduce health behavior risk factors against another intervention that only involved behavioral risk factor reduction promotion (112). The quit rates among the blue-collar workers in the integrated program were similar to quit rates for white-collar workers, and were more than twice the quit rate among blue-collar workers in the health-promotion-only program (112).

This integrated occupational/behavioral risk factor intervention approach has been applied to a number of other workplace settings, including to those that do not have a traditional building-based structure. For example, Sorensen and colleagues developed an integrated intervention to address occupational hazards, tobacco use, and weight management among motor freight workers (135). The importance of intervening in this setting is underscored by the fact that, at baseline, 40% of workers were current tobacco users and 88% were overweight or obese. The telephone-delivered intervention focused on tobacco use cessation and weight management through a social contextual framework that addressed concerns about the work environment (e.g., hazardous working conditions, stresses in meeting deadlines, limited food and physical activity options). Ten months post baseline, tobacco users who participated in the intervention were more likely to have quit using tobacco than nonparticipants: 23.8% vs. 9.1% ($p = 0.02$), although there was no change in weight management.

An integrated model in which workplaces have the same people, processes, and policies working together on improving the overall health of workers—instead of one that separates health promotion, occupational health, and employee assistant programs—is now the prevailing approach for workplace health promotion (136). Interestingly, this model has been embraced by researchers and practitioners who are concerned with eliminating exposure to occupational hazards (137, 138). In fact, the National Institute for Occupational Safety and Health adopted this approach as a research to practice priority called the Total Worker Health Program (139).

FAITH-BASED INTERVENTIONS

A number of studies have shown that faith-based settings are a feasible and acceptable venue in which to provide health information; the majority of these studies have been conducted

in African American churches (140). Although recent data suggest an increasing trend in the number of Americans who are nonreligious, a significant number (85%) of Americans report identifying with a religion, and places of worship play an important role in family life (141, 142). In particular, places of worship play a prominent role in many minority and lower-income communities and represent a trusted, credible institution that addresses both spiritual and physical health. Many places of worship espouse a social justice mission and play a key role in addressing social structural factors that influence health (e.g., food insufficiency). They also typically have structural resources that facilitate conduct of health promotion (e.g., buildings with kitchens and meeting rooms, access to groups that convene regularly); opportunities to implement policies that support health (e.g., types of foods served at church events); and use of facilities for unhealthful (e.g., smoking) and healthful (e.g., physical activity) behaviors. Intervention targets have included behavioral risk factors such as tobacco use, diet, physical activity, and cancer screening (143).

An excellent example of a church-based intervention that addresses social context is the Body and Soul program (144). The intervention included: church-wide nutrition activities and events; self-help materials that included a cookbook and nutrition video; at least one policy or environmental change (such as serving fruit and vegetables at events with food); and peer counseling calls using motivational interviewing by trained lay church members (144). Fifteen churches from six states were randomized to either the intervention or a comparison group. At the 6-month follow-up, participants in the intervention group reported significantly greater consumption of fruit and vegetables—a difference of 1.4 servings per day. Demonstrating the importance of addressing both individual- and social-level factors, a mediation analysis suggested significant mediation of the intervention effect by social support and self-efficacy, which explained 25% of the variance in dietary improvement (145). There has been increasing emphasis on the dissemination of Body and Soul outside of the research context, and on the role of Latino and Haitian churches in promoting health (146–148).

FAMILY-BASED INTERVENTIONS

The family is widely recognized as an important social context for health behavior change. Diet and physical activity behaviors, for example, emerge early in life within the family context, are heavily shaped by family members (149–151), and show stability over time (152–154). Smoking initiation is also greatly influenced by the family through parents' smoking attitudes and behaviors, parenting practices, and family functioning and cohesion (155, 156).

As a result of their fundamental role in shaping child and adolescent behaviors at crucial points in development, families offer distinct advantages as an intervention channel over many other channels. The intensity and longevity of interactions with family members result in emotional bonds that are not easily established in other settings. Such bonds make parents and family members powerful agents of change (157, 158). Moreover, parents act as gatekeepers to other contexts or channels and as a result can impact the individual through multiple channels. Finally, given that most individuals maintain ties with their family of origin across the lifecourse, families and family members have the potential to effect change over an extended time period.

As with interventions implemented through other channels, family interventions have evolved extensively over the past three decades. In early family interventions, parents were schooled in "just say no" approaches or simply provided information about the dangers of particular behaviors. In contrast, effective contemporary family interventions are comprehensive (spanning multiple settings or channels), grounded in theory, tailored to a child's developmental stage, socioculturally relevant to families, and built on research from family and prevention science (159).

A recent example is provided by Communities for Healthy Living (CHL), a family-centered obesity intervention targeting low-income preschool-aged children enrolled in Head Start and their families (160). Utilizing CBPR principles and grounded in family and empowerment theories, CHL was developed in collaboration with low-income parents, Head Start staff, and community members (69, 161, 162). The resulting intervention integrated a health communication campaign to dispel parental myths about childhood obesity; nutrition counseling sessions in the context of Head Start family outreach events; revisions to Head Start body mass index reporting procedures; and a parent-led program targeting life challenges beyond those typically addressed in obesity interventions through skill-building in conflict resolution, effective communication, and resource empowerment. Findings from a year-long assessment of CHL provide preliminary evidence supporting positive program effects on children's obesity prevalence, dietary intake, and physical activity, and parent resource empowerment and food, physical activity, and media parenting practices (160).

TRANSLATION OF CLINICAL INTERVENTIONS TO COMMUNITY SETTINGS

One of the most important studies of the impact of behavioral interventions on disease prevention was the Diabetes Prevention Program (DPP), in which diabetes *incidence* in a diverse sample of high-risk adults was reduced by 58% with intensive lifestyle intervention—about double that achieved with metformin (31%)—compared with placebo. The superior efficacy of the lifestyle intervention was maintained at the 10-year follow-up, with diabetes incidence reduced by 34% in the lifestyle group and 18% in the metformin group, compared with placebo. This is a stunning result that one might expect would lead to significant enthusiasm for behavioral and lifestyle approaches. Unfortunately, this has largely not been the case, in part because of the difficulty of scaling up the intensive lifestyle intervention. However, there are a growing number of efforts to adapt the DPP intervention to community settings, using more scalable approaches that also take contextual factors into consideration. Marrero and colleagues tested the feasibility and effectiveness of an adapted version of the DPP intervention delivered in a group format through YMCAs (163, 164). "Lay leaders" were employees in the YMCA. Participants were randomized to the group intervention or to brief counseling with diabetes risk screening. The intervention addressed access to physical activity resources and contextual factors through the group process. Those in the group intervention achieved a 6% weight loss that was maintained at 6 and 14 months post intervention; a significant reduction in total cholesterol was seen over the same time period. The intervention delivery cost was $205 per participant, compared with $1,476 in the original DPP study.

Another example of utilizing a community-based approach for addressing health behaviors is the use of peer supporters. Peer-delivered interventions take many forms, and have been evaluated primarily in the context of management of chronic disease and follow-up of abnormal screening (165–169). Peer interventions have been delivered in a variety of ways (e.g., patient navigators, community health workers, peer supporters), including use of peers to implement behavior change activities (91) and to provide individualized support for management of chronic diseases. A common feature, however, is the use of peers to contextualize the health issue and to provide strategies for integrating behavior change into the realities of one's daily life. For example, Emmons et al. utilized childhood cancer survivors to deliver a peer-based intervention for smoking cessation (91). By integrating a focus on the childhood cancer experience, and by considering contextual factors facing survivors who smoke (e.g., lower income, higher levels of depression, comorbidities), the intervention led to a doubling of the cessation rate compared to those receiving a self-help intervention.

An excellent example of a peer-based approach that has been tested in global health settings is the Peers for Progress program, which focuses on peer-delivered approaches for diabetes self-management and has been implemented in Thailand and three sub-Saharan African countries (168). This is also a good example of the importance of addressing both social context and behavioral skills. Self-management is essential to reducing the risks of diabetes, and it is unlikely that policy or social-level interventions alone could replace the need for self-management activities. Peers can provide the kind of ongoing support that is needed for sustained self-management of diabetes, and for addressing social contextual factors that frequently serve as barriers to lasting change. In the Peers for Progress program, peers were nonprofessionals who had diabetes or close familiarity with its management, and were integrated with the healthcare system. Peers' key functions included providing assistance in daily management of diabetes; social and emotional support to encourage the management of behavior and to help patients cope with negative emotions; linkage to clinical care and community resources; and ongoing support. A recent evaluation found good evidence for significant improvement in diabetes management (e.g., reduction of HbA1c levels, blood pressure, and/or weight), and quality of life outcomes (168). Also, there were positive outcomes related to the long-term sustainability of the interventions, which is often a significant problem for behavioral interventions in a variety of settings. Fisher and colleagues note that the demonstrated feasibility of peer support in the global health setting suggests that the strategy could be applied to the patient-centered medical home in the United States (168).

POLICY APPROACHES TO BEHAVIOR CHANGE

As discussed in Chapters 12 and 15 of this book, efforts to change behaviors and the health outcomes associated with them have also involved policy interventions. The Massachusetts farmers' market coupon program for low-income elders is an excellent example of a policy intervention designed to increase individual fruit and vegetable consumption by increasing access (170, 171). Through collaboration between the Massachusetts Department of Public Health and the Department of Food and Agriculture, farmers' market coupons were distributed through

elderly nutrition projects throughout the state. In 1992, almost $86,000 in coupons were distributed by 23 agencies to 17,200 older adults; 73% of the coupons were redeemed, and 32% of the seniors reported buying significantly more fruit and vegetables since receiving the coupons (171). Coupons distributed through this program brought an additional $62,000 in revenue to the markets, in addition to money spent at the markets after the coupons had been spent. This is an excellent example of an access-oriented intervention that targeted both individual behavior and organizational- and policy-level change, building on an interagency collaboration that addressed separate and overlapping goals of each agency.

The importance of considering the behavioral intervention evidence base in policy also needs to be recognized. As discussed in Chapter 15, policy interventions, especially those that address the social determinants of health, invariably require engagement with the political process and with powerbrokers who shape policies in different governmental entities. Although policy interventions hold significant promise to influence population health, if they are not evidence based, they will likely have little impact. Eyler and colleagues examined whether evidence related to children's participation in physical education (PE) is reaching policymakers and being incorporated into legislation (172). Although over the last 10 years there has been a substantial increase in the number of PE bills introduced at the state level, doubling from about 70 bills in 2001 to 140 in 2007, only about 25% of the bills that were introduced contained evidence-based elements. Furthermore, only 21% of those introduced were put into law, and just one-third of those bills contained one or more evidence-based elements. There is significant potential for public policy to drive access to evidence-based interventions in an equalizing manner—across all social gradients. However, this has been a slow and limited approach, with very little emphasis on evaluating the impact of policy on behavioral outcomes. Eyler also found that only 35% of the policies enacted required evaluation of the policy after enactment. In general there is very little focus on public policy research, particularly in the public health and behavioral medicine arenas. This is a critical area for increased engagement from the behavioral science community, and one that will likely pay dividends in terms of increasing both translation of our evidence base and our relevance.

THE POTENTIAL FOR INTEGRATED POLICY AND BEHAVIORAL INTERVENTIONS TO INCREASE BEHAVIOR CHANGE IMPACT

The interaction between policy and behavioral interventions is often overlooked, but is a critical pathway for improving population health, especially in the context of socially patterned behaviors. An excellent example is the implementation of comprehensive tobacco control through the Massachusetts Medicaid program, MassHealth. In the early 1990s, the smoking prevalence in Massachusetts was slightly higher than that of the US (173, 174). In response, in 1993 the state established the Massachusetts Tobacco Control Program (MTCP); this was one of the few truly comprehensive tobacco control programs in the United States at that time. Informed by evidence on social determinants of health and the importance of a population approach, the MTCP embraced a population approach with a strong social determinant framework that focused on policy and environmental changes, including establishment of community-based services to promote anti-tobacco policies and services (175, 176). Even when compared with nationwide downward

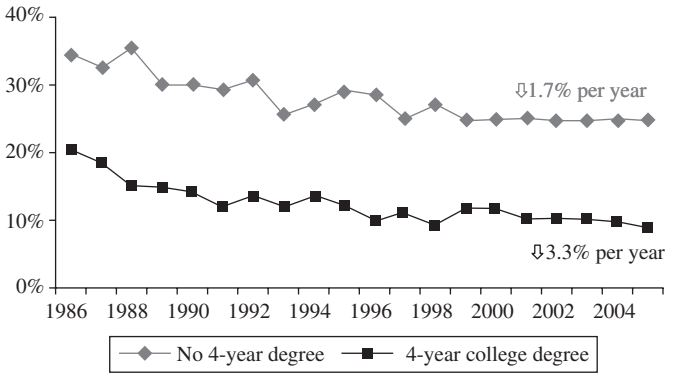


FIGURE 10.3: Adult Smoking Prevalence by Educational Level, Massachusetts, 1986–2005.

trends, Massachusetts experienced a significant decrease in tobacco use that was directly tied to the MTCP (175–177).

However, the overall smoking prevalence trends in Massachusetts masked significant disparities that were particularly stark by level of education (see Figure 10.3). From 1986 to 2005, the prevalence of smoking for those with a college education had decreased by about 3.3% annually, but for those without a college education, the decline rate was about half of that (1.7%). This picture changed drastically in 2006 when Massachusetts became one of 22 US states whose Medicaid program included tobacco cessation treatment with coverage for pharmacotherapy, and one of only 6 states to additionally cover behavioral smoking cessation counseling for Medicaid members (173, 174). In the 2.5 years after the policy change, over 75,000 Medicaid smokers took advantage of the benefit—about 40% of eligible smokers, leading to a smoking prevalence decrease from 38.3% to 28.8% among subscribers (173, 174). Although the rate of quit attempts did not change, the *success rate* did—jumping from 6.6% to 19% after the benefits were added (173, 174). Most importantly, these reductions in smoking were associated with 46% and 49% reductions in myocardial infarction and coronary atherosclerosis admissions, respectively, among the state Medicaid beneficiaries (173).

In the case of Massachusetts, it was a policy change targeting access to behavioral cessation interventions that led to changes in smoking prevalence for a disadvantaged population. Oftentimes, discussions of behavioral interventions are focused on what Albert Bandura refers to as a "contentious dualism" of individualist versus structuralist approaches to health (70). In this dichotomy, behavioral counseling would be the individualistic approach while the structuralist approach would be a policy change. We would argue that this is a false dichotomy, and that it is critical that public health begin to address this dualism of individualist versus structuralist approaches to intervention design and delivery.

THE FUTURE OF BEHAVIOR CHANGE INTERVENTIONS: WHERE TO FROM HERE?

If we are to expedite the improvement of population health, we must look to all areas of science for strategies that could be leveraged to improve health behaviors. The National Human Genome

Research Institute's research priorities for the coming decade include the recommendation to conduct research that informs the use of genomic information to improve behavior change interventions (178). The rapidly emerging science of epigenetics is demonstrating that the very social and physical environments that increase health behavior risks in some population groups may modify the genome to produce new phenotypes (179). McBride and colleagues note that one pathway by which genetics may help to improve response to behavioral interventions relates to adherence (180). That is, individuals vary in how they respond both physically and emotionally to behavioral recommendations. They note that there is methodologically limited but growing evidence that genetic variation accounts for some of the differences in physiological responses to caloric restriction, dietary composition, and engagement in moderate or intensive physical activity regimens. For example, some people may have a genetic variant that increases adverse physical reactions to vigorous exercise, and thus may lead to higher levels of sedentary behavior. Although one might consider these people "nonadherent" to health recommendations, their unwillingness to exercise may be a functional response. Thus, for these individuals, a better approach to increasing physical activity may be to target increasing moderate activities such as walking, or focusing on dietary strategies to maintain or lose weight.

Genetic risk information is also being used to activate emotional processes in response to behavior change recommendations. For example, Hay and colleagues used prototypic genetic risk feedback with first-degree relatives of patients with melanoma to increase sun protection and skin screening (181). Risk level was associated with higher perceived risk and behavioral intentions for sun protection and skin screening. This study highlights the importance of building an evidence base to inform genomic risk communication practice, and is just one example of many that are beginning to explore these domains. McBride et al. propose several steps to initiate research on genomics-informed customization of behavior change interventions that are relevant to population health, two of which are particularly relevant to considering the social context of behavior change: *(1)* review the scientific literature to identify areas with potential to influence behavioral adherence (e.g., phenotypes that are likely to have strong genetic underpinnings), especially those that are common across behaviors (e.g., dopaminergic rewards associated with eating and brain-derived neurotrophic factors involved in energy metabolism); and *(2)* develop conceptual models to map the interrelationships of relevant biological, psychological, and macro-level factors that influence adherence. They also recommend conducting prospective comparative effectiveness studies to evaluate whether genotype-informed customization adds value, and doing so while collecting data on—and addressing—social contextual factors, in order to be maximally informative. Although it is easy to eschew genetics as being unlikely to have much explanatory power related to health behavior and health outcomes, a growing body of research suggests that this would be unwise, and that significant innovation may come from considering the genetic context alongside the social and physical context.

CONCLUSION

This chapter has briefly summarized the evolution and contributions of behavioral interventions for health behavior change, and has illustrated the key role that social context plays in health

behavior. As this chapter has tried to illustrate, there are now many examples of behavioral interventions that address social contextual factors. Increasingly, interventions take a population-health approach, are designed to address a range of motivational levels, and are pro-actively focused on primary prevention. The key points of the chapter are below.

1. Health behaviors are undertaken by individuals, but they are shaped and can be modified by intrapersonal, familial, historical, social, political, and other events that occur inside and outside the individual.
2. The most successful interventions are multilevel interventions that consider and address all levels of influence on health behaviors.
3. Public health has historically been plagued by an unhelpful contentious dualism that demarcates interventions into individual and structural approaches; however, both approaches are often needed to improve population health.
4. The most effective theories and approaches for improving population health will arise from meaningful interdisciplinary collaborations.
5. The evidence base for multilevel interventions is growing, and attention needs to be shifted toward how to scale up effective interventions.

Much attention has been paid to the limits of behavior change interventions and the small effect sizes of large-scale community-based behavior change efforts. A common concern of early interventions was that much of the onus for change was placed on individuals without acknowledging the strong impact that the social context exerted on behaviors (71, 75). In a 1986 essay titled the *Tyranny of Health Promotion*, Marshall Becker, one of the developers of the health belief model and a key figure in health promotion, advocated for a shift toward behavior change interventions that paid equal attention to health education and social determinants of health (71). His and others' calls for more interventions that are not solely focused on buffering psychological processes have led to significant changes in both the design of individually targeted interventions and the expansion of targets of intervention activities.

As clear as it is that effective health promotion interventions can no longer ignore social contextual factors, we still argue that efforts to address the broader social context should not completely abandon efforts to intervene at the individual level. As Altman states, the key point in prevention research is to identify the web of causation and to intervene on as many levels as possible in that web (182). The most effective intervention strategies are likely to incorporate both the individual whose health behavior is in question and the larger community and governmental forces that influence the life of that individual. Furthermore, studies that integrate individual interventions with larger systems intervention strategies may result in methodological interventions that will improve our understanding of how best to conceptualize, intervene upon, and assess health behavior change efforts. The very best approaches, in our opinion, focus on multiple levels of influence, and develop collaborations with community- and policy-based channels to address health behaviors in the context of social determinants and access issues. The approach to tobacco control adopted in Massachusetts provides an excellent example of the impact of these approaches. Such upstream approaches that are a central part of efforts to make an impact at the population level require coordinated, sustained effort, and advocacy for policy and legislative changes at the local, state, and national levels. This is particularly needed in efforts to address obesity epidemics among

children and adults in developed and developing nations. Population approaches that address the cheaper and wider availability and aggressive marketing of unhealthy foods are needed as well as interventions that target those vulnerable populations specifically.

Adopting either a solely individual, biomedical, or social epidemiology perspective limits the abilities to impact population health. Although an extensive social epidemiological literature addresses the relationship between social factors and health outcomes, little attention has been paid to the science by which social factors can be changed and how to change them to improve population health (107). A major drawback in much of health-related science has been the willingness to put significant stock in single-level solutions—be it drugs, Web-based interventions, or health policy. It is highly unlikely that any such solution alone will truly impact population health. However, policy interventions that target access to behavioral interventions can definitely lead to changes in health behaviors among disadvantaged populations. As noted earlier, public health has historically been plagued by a contentious dualism pitting behavioral intervention approaches against structural approaches recommended by social epidemiology. In our opinion, this is a false dichotomy. We have argued that, at best, this is unhelpful and, at worst, destructive. Partnerships between behavioral scientists and social epidemiologists that focus on solution-oriented research are critical toward bridging the population health gap (67, 107), and will go a long way toward increasing the relevance of both areas of science. In our view, the most successful interventions are multilevel, considering and addressing the multiple levels of influence on health behaviors, including the whole range from intrapersonal factors to the community and policy context. The field is increasingly moving toward integrating interventions across levels, which is critical. It is unlikely that any approach focused on a single set of theories or tools will yield the level of behavior change needed to eliminate either health disparities or preventable diseases at the population level. Meaningful interdisciplinary collaboration is likely to yield the most effective approaches, as examples in this chapter have illustrated. A key feature that we all must attend to is scale—how to design and evaluate interventions so that, if effective, the findings are generalizable and can be applied at the population level to benefit the health of all.

REFERENCES

1. Fisher EB, Fitzgibbon ML, Glasgow RE, Haire-Joshu D, Hayman LL, Kaplan RM, et al. Behavior matters. Am J Prev Med. 2011;40(5):e15–e30.
2. Alwan A, MacLean DR, Riley LM, d'Espaignet ET, Mathers CD, Stevens GA, et al. Monitoring and surveillance of chronic non-communicable diseases: progress and capacity in high-burden countries. Lancet. 2010;376(9755):1861–8.
3. Danaei G, Ding EL, Mozaffarian D, Taylor B, Rehm J, Murray CJ, et al. The preventable causes of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. PLoS Medicine. 2009;6(4):e1000058.
4. Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Actual causes of death in the United States, 2000. JAMA. 2004;291(10):1238–45.
5. Eheman C, Henley SJ, Ballard Barbash R, Jacobs EJ, Schymura MJ, Noone AM, et al. Annual report to the nation on the status of cancer, 1975–2008, featuring cancers associated with excess weight and lack of sufficient physical activity. Cancer. 2012;118(9):2338–66.

6. Winzer BM, Whiteman DC, Reeves MM, Paratz JD. Physical activity and cancer prevention: a systematic review of clinical trials. Cancer Cause Control. 2011;22(6):811–26.
7. Hoyert D, Xu J. Deaths: preliminary data for 2011. National Vital Statistics Reports. 2012;61(6).
8. Koh HK, Sebelius KG. Promoting prevention through the affordable care act. N Engl J Med. 2010;363(14):1296–9.
9. Villerme L. A description of the physical and moral state of workers employed in cotton, wool and silk mills. In: Buck C, Llopis A, Najera E, Terris M, editors. The challenge of epidemiology: issues and selected readings. Washington, DC: PAHO/WHO 1988; 1840. pp. 33–6.
10. Agaku I, King B, Dube SR. Current cigarette smoking among adults—United States, 2011. Morb Mortal Wkly Rep. 2012;61:889–94.
11. Haenszel W, Shimkin MB, Miller HP. Tobacco smoking patterns in the United States. Public Health Monogr. 1956(45):1–105.
12. Syamlal G, Mazurek JM. Current cigarette smoking prevalence among working adults—United States, 2004–2010. Morb Mortal Wkly Rep. 2011;60(38):1305–9.
13. CDC. Trends in current cigarette smoking among high school students and adults, United States, 1965–2011. 2013 (Accessed August 12, 2013 from http://www.cdc.gov/tobacco/data_statistics/tables/trends/cig_smoking/).
14. Prochaska JJ. Smoking and mental illness: breaking the link. N Engl J Med. 2011;365(3):196–8.
15. Banham L, Gilbody S. Smoking cessation in severe mental illness: what works? Addiction. 2010;105(7):1176–89.
16. Barbeau EM, Krieger N, Soobader MJ. Working class matters: socioeconomic disadvantage, race/ethnicity, gender, and smoking in NHIS 2000. Am J Public Health. 2004 Feb;94(2):269–78.
17. Schroeder S. Stranded in the periphery: the increasing marginalization of smokers. N Engl J Med. 2008 May 22;358(21):2284–6.
18. Margerison-Zilko C, Cubbin C. Socioeconomic disparities in tobacco-related health outcomes across racial/ethnic groups in the United States: National Health Interview Survey 2010. Nicotine Tob Res. 2013;15(6):1161–5.
19. Okechukwu C, Bacic J, Cheng KW, Catalano R. Smoking among construction workers: the nonlinear influence of the economy, cigarette prices, and antismoking sentiment. Soc Sci Med. 2012;75(8):1379–86.
20. Gallus S, Ghislandi S, Muttarak R, Bosetti C. Effects of the economic crisis on smoking prevalence and number of smokers in the USA. Tob Control. 2013.
21. Shavers VL, Lawrence D, Fagan P, Gibson JT. Racial/ethnic variation in cigarette smoking among the civilian US population by occupation and industry, TUS-CPS 1998–1999. Prev Med. 2005;41(2):597–606.
22. Lee DJ, Fleming LE, Arheart KL, LeBlanc WG, Caban AJ, Chung-Bridges K, et al. Smoking rate trends in U.S. occupational groups: the 1987 to 2004 National Health Interview Survey. J Occup Environ Med. 2007;49(1):75–81.
23. Ham DC, Przybeck T, Strickland JR, Luke DA, Bierut LJ, Evanoff BA. Occupation and workplace policies predict smoking behaviors: analysis of national data from the current population survey. J Occup Environ Med. 2011;53(11):1337–45.
24. Fujishiro K, Stukovsky KD, Roux AD, Landsbergis P, Burchfiel C. Occupational gradients in smoking behavior and exposure to workplace environmental tobacco smoke: the multi-ethnic study of atherosclerosis. J Occup Environ Med. 2012;54(2):136–45.
25. de Castro AB, Garcia G, Gee GC, Tsai JH, Rue T, Takeuchi DT. Smoking and the Asian American workforce in the National Latino and Asian American Study. Am J Ind Med. 2010;53(2):171–8.
26. Bang KM, Kim JH. Prevalence of cigarette smoking by occupation and industry in the United States. Am J Ind Med. 2001;40(3):233–9.

27. Flegal KM, Carroll MD, Kit BK, Ogden CL. Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999–2010. JAMA. 2012;307(5):491–7.

28. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of obesity and trends in body mass index among US children and adolescents, 1999–2010. JAMA. 2012;307(5):483–90.

29. Kruger J, Yore M, Solera M, Moeti R. Prevalence of fruit and vegetable consumption and physical activity by race/ethnicity—United States, 2005. Morb Mortal Wkly Rep. 2007;56(13):301–4.

30. USDA. Dietary guidelines for Americans, 2005. Washington, DC: US Government Printing Office; 2005.

31. Rossen L, Schoendorf K. Measuring health disparities: trends in racial-ethnic and socioeconomic disparities in obesity among 2 to 18 year old youth in the United States, 2001–2010. Ann Epidemiol. 2012;22(10): 698–704.

32. Kirkpatrick SI, Dodd KW, Reedy J, Krebs-Smith SM. Income and race/ethnicity are associated with adherence to food-based dietary guidance among US adults and children. J. Acad. Nutr. Diet. 2012;112(5):624–35 e6.

33. Grimm KA, Foltz JL, Blanck HM, Scanlon KS. Household income disparities in fruit and vegetable consumption by state and territory: results of the 2009 Behavioral Risk Factor Surveillance System. J. Acad. Nutr. Diet. 2012;112(12):2014–21.

34. Ham SA, Ainsworth BE. Disparities in data on Healthy People 2010 physical activity objectives collected by accelerometry and self-report. Am J Public Health. 2010 Apr 1;100 Suppl 1:S263–8.

35. Whitt-Glover MC, Taylor WC, Floyd MF, Yore MM, Yancey AK, Matthews CE. Disparities in physical activity and sedentary behaviors among US children and adolescents: prevalence, correlates, and intervention implications. J Public Health Pol. 2009;30(Suppl 1):S309–34.

36. Lloyd-Jones D, Adams RJ, Brown TM, Carnethon M, Dai S, De Simone G, et al. Heart disease and stroke statistics—2010 update: a report from the American Heart Association. Circulation. 2010;121(7):e46-e215.

37. Troutt DD. Thin red line: how the poor still pay more. San Francisco, CA: Consumers Union of the US, West Coast Regional Office. 1993.

38. Chung C, Myers SL. Do the poor pay more for food? An analysis of grocery store availability and food price disparities. J Consum Aff. 1999;33(2):276–96.

39. Moore LV, Diez Roux AV. Associations of neighborhood characteristics with the location and type of food stores. Am J Public Health. 2006;96(2):325–31.

40. Hackbarth DP, Silvestri B, Cosper W. Tobacco and alcohol billboards in 50 Chicago neighborhoods: market segmentation to sell dangerous products to the poor. J Public Health Pol. 1995:213–30.

41. Luke D, Esmundo E, Bloom Y. Smoke signs: patterns of tobacco billboard advertising in a metropolitan region. Tob Control. 2000;9(1):16–23.

42. Barbeau EM, Wolin KY, Naumova EN, Balbach E. Tobacco advertising in communities: associations with race and class. Prev Med. 2005;40(1):16–22.

43. Smoyer-Tomic KE, Spence JC, Raine KD, Amrhein C, Cameron N, Yasenovskiy V, et al. The association between neighborhood socioeconomic status and exposure to supermarkets and fast food outlets. Health Place. 2008;14(4):740–54.

44. Kwate NOA. Fried chicken and fresh apples: racial segregation as a fundamental cause of fast food density in black neighborhoods. Health Place. 2008;14(1):32–44.

45. Kwate NOA, Yau C-Y, Loh J-M, Williams D. Inequality in obesigenic environments: Fast food density in New York City. Health Place. 2009;15(1):364–73.

46. Gerlach KK, Shopland DR, Hartman AM, Gibson JT, Pechacek TF. Workplace smoking policies in the United States: results from a national survey of more than 100,000 workers. Tob Control. 1997;6(3):199–206.

47. Brownson RC, Hopkins DP, Wakefield MA. Effects of smoking restrictions in the workplace. Annu Rev Public Health. 2002;23:333–48.

48. Alexander LA, Crawford T, Mendiondo MS. Occupational status, work-site cessation programs and policies and menthol smoking on quitting behaviors of US smokers. Addiction. 2010;105(Suppl 1):95–104.

49. Kaper J, Wagena E, Willemsen M, Van Schayck C. Reimbursement for smoking cessation treatment may double the abstinence rate: results of a randomized trial. Addiction. 2005;100(7):1012–20.

50. Reda AA, Kaper J, Fikrelter H, Severens JL, van Schayck CP. Healthcare financing systems for increasing the use of tobacco dependence treatment. Cochrane Database Syst Rev. 2009;2.

51. Barbeau EM, Li YI, Sorensen G, Conlan KM, Youngstrom R, Emmons K. Coverage of smoking cessation treatment by union health and welfare funds. Am J Public Health. 2001;91(9):1412–5.

52. Sorensen G, Quintiliani L, Pereira L, Yang M, Stoddard A. Work experiences and tobacco use: findings from the Gear Up for Health Study. J Occup Environ Med. 2009;51(1):87–94.

53. Abrams DB, Biener L. Motivational characteristics of smokers at the workplace: a public health challenge. Prev Med. 1992;21(6):679–87.

54. Sorensen G, Emmons K, Stoddard AM, Linnan L, Avrunin J. Do social influences contribute to occupational differences in quitting smoking and attitudes toward quitting? Am J Health Promot. 2002;16(3):135–41.

55. Okechukwu C, Krieger N, Sorensen G, Li Y, Barbeau EM. Testing hypothesized psychosocial mediators: lessons learned in the MassBUILT study. Health Educ Behav. 2011;38(4):404–11.

56. Okechukwu C, Dutra L, Bacic J, El Ayadi A, Emmons K. Home matters: work and household predictors of smoking and cessation among blue-collar workers. Prev Med. 2012;56(2):130–4.

57. Okechukwu C, Nguyen K, Hickman NJ. Partner smoking characteristics: associations with smoking and quitting among blue-collar apprentices. Am J Ind Med. 2010;53(11):1102–8.

58. Graham H. Smoking prevalence among women in the European Community 1950–1990. Soc Sci Med. 1996;43(2):243–54.

59. Graham H. When life's a drag: women, smoking and disadvantage. London: Her Majesty's Stationery Office; 1993.

60. Lundberg U. Stress responses in low-status jobs and their relationship to health risks: musculoskeletal disorders. Ann N Y Acad Sci. 1999;896:162–72.

61. Peretti-Watel P, Constance J. "It's all we got left": why poor smokers are less sensitive to cigarette price increases. Int J Environ Res Public Health. 2009;6(2):608–21.

62. Quinn MM, Sembajwe G, Stoddard AM, Kriebel D, Krieger N, Sorensen G, et al. Social disparities in the burden of occupational exposures: results of a cross-sectional study. Am J Ind Med. 2007;50(12):861–75.

63. Krieger N, Waterman PD, Hartman C, Bates LM, Stoddard AM, Quinn MM, et al. Social hazards on the job: workplace abuse, sexual harassment, and racial discrimination—a study of Black, Latino, and White low-income women and men workers in the United States. Int J Health Serv. 2006;36(1):51–85.

64. Krieger N, Kaddour A, Koenen K, Kosheleva A, Chen JT, Waterman PD, et al. Occupational, social, and relationship hazards and psychological distress among low-income workers: implications of the "inverse hazard law." J Epidemiol Community Health. 2011;65(3):260–72.

65. Okechukwu C, Souza K, Davis KD, de Castro AB. Discrimination, harassment, abuse, and bullying in the workplace: contribution of workplace injustice to occupational health disparities. Am J Ind Med. 2013.

66. Okechukwu C, Krieger N, Chen J, Sorensen G, Li Y, Barbeau EM. The association of workplace hazards and smoking in a U.S. multiethnic working-class population. Public Health Rep. 2010;125(2):225–33.

67. Sorensen G, Emmons K, Hunt MK, Barbeau EM, Goldman R, Peterson K, et al. Model for incorporating social context in health behavior interventions: applications for cancer prevention for working-class, multiethnic populations. Prev Med. 2003;37(3):188–97.

68. Sorensen G, Barbeau EM, Hunt MK, Emmons K. Reducing social disparities in tobacco use: a social-contextual model for reducing tobacco use among blue-collar workers. Am J Public Health. 2004;94(2):230–9.

69. Davison KK, Jurkowski JM, Lawson HA. Reframing family-centred obesity prevention using the family ecological model. Public Health Nutr. 2013;16(10):1861–9.

70. Bandura A. Health promotion by social cognitive means. Health Educ Behav. 2004;31(2):143–64.

71. Becker M. The tyranny of health promotion. Public Health Rev. 1986;14(1):15.

72. Emmons KM, Rollnick S. Motivational interviewing in health care settings: opportunities and limitations. Am J Prev Med. 2001;20(1):68–74.

73. Montano DE, Kasprzyk D. Theory of reasoned action, theory of planned behavior, and the integrated behavioral model. In: Glantz K, Rimer BK, Viswanath K, editors. Health behavior and health education: Theory, research, and practice. San Francisco: Jossey-Bass; 2008. pp. 67–95.

74. Champion VL, Skinner CS. The health belief model. In: Glantz K, Rimer BK, Viswanath K, editors. Health behaviour and health education; theory, research, and practice. San Francisco: Jossey-Bass; 2008. pp. 45–65.

75. Goodson P. Theory in health promotion research and practice: Thinking outside the box: Jones & Bartlett Learning; 2009.

76. Resnicow K, Page SE. Embracing chaos and complexity: a quantum change for public health. Am J Public Health. 2008;98(8):1382–9.

77. Laws RA, St George AB, Rychetnik L, Bauman AE. Diabetes prevention research: a systematic review of external validity in lifestyle interventions. Am J Prev Med. 2012;43(2):205–14.

78. Green LW, Glasgow RE. Evaluating the relevance, generalization, and applicability of research issues in external validation and translation methodology. Eval Health Prof. 2006;29(1):126–53.

79. Rose G. Sick individuals and sick populations. Int J Epidemiol. 2001;30(3):427–32.

80. Miller WR, Rollnick S. Motivational interviewing: preparing people for change. 3rd ed. New York: Guilford Press; 2002.

81. Rollnick S, Miller WR. What is motivational interviewing? Behav Cogn Psychoth. 1995;23:325–34.

82. Colby SM, Monti PM, Barnett NP, Rohsenow DJ, Weissman K, Spirito A, et al. Brief motivational interviewing in a hospital setting for adolescent smoking: a preliminary study. J Consult Clin Psych. 1998;66(3):574.

83. Resnicow K, DiIorio C, Soet JE, Borrelli B, Hecht J, Ernst D. Motivational interviewing in health promotion: it sounds like something is changing. Health Psychol. 2002;21(5):444.

84. Hettema JE, Hendricks PS. Motivational interviewing for smoking cessation: a meta-analytic review. J Consult Clin Psych. 2010;78(6):868.

85. Rubak S, Sandbæk A, Lauritzen T, Christensen B. Motivational interviewing: a systematic review and meta-analysis. Brit J Gen Pract. 2005;55(513):305.

86. Martins RK, McNeil DW. Review of motivational interviewing in promoting health behaviors. Clin Psychol Rev. 2009;29(4):283–93.

87. Lai D, Cahill K, Qin Y, Tang JL. Motivational interviewing for smoking cessation. Cochrane Database Syst Rev. 2010;1.

88. Loewenstein G, Asch DA, Friedman JY, Melichar LA, Volpp KG. Can behavioural economics make us healthier?. 2012;344.

89. Blumenthal KJ, Saulsgiver KA, Norton L, Troxel AB, Anarella JP, Gesten FC, et al. Medicaid incentive programs to encourage healthy behavior show mixed results to date and should be studied and improved. Health Affairs. 2013;32(3):497–507.

90. Goode AD, Reeves MM, Eakin EG. Telephone-delivered interventions for physical activity and dietary behavior change: an updated systematic review. Am J Prev Med. 2012;42(1):81–8.

91. Emmons KM, Puleo E, Park E, Gritz ER, Butterfield RM, Weeks JC, et al. Peer-delivered smoking counseling for childhood cancer survivors increases rate of cessation: the partnership for health study. J Clin Oncol. 2005;23(27):6516–23.

92. Piette JD. Satisfaction with automated telephone disease management calls and its relationship to their use. Diabetes Educator. 2000;26(6):1003–10.
93. Piette JD, Weinberger M, McPhee SJ, Mah CA, Kraemer FB, Crapo LM. Do automated calls with nurse follow-up improve self-care and glycemic control among vulnerable patients with diabetes? Am J Med. 2000;108(1):20–7.
94. Anderson CM, Zhu S-H. Tobacco quitlines: looking back and looking ahead. Tob Control. 2007;16(Suppl 1):i81–i6.
95. Zhu S-H, Anderson CM, Tedeschi GJ, Rosbrook B, Johnson CE, Byrd M, et al. Evidence of real-world effectiveness of a telephone quitline for smokers. N Eng J Med. 2002;347(14):1087–93.
96. Stead LF, Perera R, Lancaster T. Telephone counselling for smoking cessation. Cochrane Database Syst Rev. 2006;3.
97. Lichtenstein E, Zhu S-H, Tedeschi GJ. Smoking cessation quitlines: an underrecognized intervention success story. Am Psychol. 2010;65(4):252.
98. Fiore MC, Croyle RT, Curry SJ, Cutler CM, Davis RM, Gordon C, et al. Preventing 3 million premature deaths and helping 5 million smokers quit: a national action plan for tobacco cessation. Am J Public Health. 2004;94(2):205–10.
99. Norman GJ, Zabinski MF, Adams MA, Rosenberg DE, Yaroch AL, Atienza AA. A review of eHealth interventions for physical activity and dietary behavior change. Am J Prev Med. 2007;33(4):336–45. e16.
100. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. Epidemiol Rev. 2010;32(1):56–69.
101. Smith A. Nearly half of American adults are smartphone owners. Pew Internet & American Life Project. 2012. Available at: http://www.pewinternet.org/Reports/2012/Smartphone-Update-2012/Findings.aspx (Accessed September 22, 2013).
102. Viswanath K, Nagler RH, Bigman-Galimore CA, McCauley MP, Jung M, Ramanadhan S. The communications revolution and health inequalities in the 21st century: implications for cancer control. Cancer Epidem Biomar Prev. 2012;21(10):1701–8.
103. Viswanath K. Cyberinfrastructure: an extraordinary opportunity to bridge health and communication inequalities? Am J Prev Med. 2011;40(5):S245–S8.
104. Kahn JG, Yang JS, Kahn JS. "Mobile" health needs and opportunities in developing countries. Health Affair. 2010;29(2):252–8.
105. Sorensen G, Emmons K, Hunt MK, Johnston D. Implications of the results of community intervention trials. Annu Rev Publ Health. 1998;19(1):379–416.
106. Merzel C, D'afflitti J. Reconsidering community-based health promotion: promise, performance, and potential. Am J Public Health. 2003;93(4):557–74.
107. Wallerstein NB, Yen IH, Syme SL. Integration of social epidemiology and community-engaged interventions to improve health equity. Am J Public Health. 2011;101(5):822–30.
108. Israel BA, Eng E, Schulz AJ, Parker EA. Introduction to methods in community-based participatory research for health. In: Israel BA, Eng E, Schulz AJ, Parker EA, editors. Methods in Community-Based Participatory Research. 2005. San Francisco, CA: Jossey-Bass; 2005.
109. Minkler M, Blackwell AG, Thompson M, Tamir H. Community-based participatory research: implications for public health funding. Am J Public Health. 2003;93(8):1210–3.
110. Berge JM, Mendenhall TJ, Doherty WJ. Using community-based participatory research (CBPR) to target health disparities in families. Fam Relat. 2009;58(4):475–88.
111. Campbell MK, Resnicow K, Carr C, Wang T, Williams A. Process evaluation of an effective church-based diet intervention: body and Soul. Health Educ Behav. 2007;34(6):864–80.

112. Sorensen G, Stoddard AM, LaMontagne AD, Emmons K, Hunt MK, Youngstrom R, et al. A comprehensive worksite cancer prevention intervention: behavior change results from a randomized controlled trial (United States). Cancer Cause Control. 2002;13(6):493–502.

113. Cahill K, Moher M, Lancaster T. Workplace interventions for smoking cessation. Cochrane Database Syst Rev. 2008;4(4).

114. Moher M, Hey K, Lancaster T. Workplace interventions for smoking cessation. Cochrane Database Syst Rev. 2005(2):CD003440.

115. Okechukwu C, Krieger N, Sorensen G, Yi L, Barbeau EM. Massbuilt: effectiveness of an apprenticeship site-based smoking cessation intervention for unionized building trades workers. Cancer Cause Control. 2009;20(6):887–94.

116. Sorensen G, Linnan L, Hunt MK. Worksite-based research and initiatives to increase fruit and vegetable consumption. Prev Med. 2004;39:94–100.

117. Quintiliani L, Sattelmair J, Sorensen G. The workplace as a setting for interventions to improve diet and promote physical activity. Documento técnico preparado para el evento conjunto OMS/Foro Económico Mundial sobre la prevención de las enfermedades no transmisibles en el lugar de trabajo]. Ginebra: Organización Mundial de la Salud; 2007.

118. Mhurchu CN, Aston LM, Jebb SA. Effects of worksite health promotion interventions on employee diets: a systematic review. BMC Public Health. 2010;10(1):62.

119. Marcus BH, Emmons KM, Simkin-Silverman LR, Linnan LA, Taylor ER, Bock BC, et al. Evaluation of motivationally tailored vs. standard self-help physical activity interventions at the workplace. Am J Health Promot. 1998;12(4):246–53.

120. Conn VS, Hafdahl AR, Cooper PS, Brown LM, Lusk SL. Meta-analysis of workplace physical activity interventions. Am J Prev Med. 2009;37(4):330–9.

121. Yancey AK, McCarthy WJ, Taylor WC, Merlo A, Gewa C, Weber MD, et al. The Los Angeles Lift Off: a sociocultural environmental change intervention to integrate physical activity into the workplace. Prev Med. 2004;38(6):848–56.

122. Hammer LB, Kossek EE, Anger WK, Bodner T, Zimmerman KL. Clarifying work–family intervention processes: The roles of work–family conflict and family-supportive supervisor behaviors. J Appl Psychol. 2011;96(1):134.

123. Secret M, Sprang G. The effects of family-friendly workplace environments on work-family stress of employed parents. J Soc Serv Res. 2002;28(2):21–45.

124. Richmond R, Kehoe L, Heather N, Wodak A. Evaluation of a workplace brief intervention for excessive alcohol consumption: the workscreen project. Prev Med. 2000;30(1):51–63.

125. Anderson BK, Larimer ME. Problem drinking and the workplace: An individualized approach to prevention. Psychol Addict Behav. 2002;16(3):243.

126. Nerin I, Crucelaegui A, Más A, Villalba JA, Guillén D, Gracia A. Results of a comprehensive workplace program for the prevention and treatment of smoking addiction. Arch Bronconeumol. 2005;41(4):197–201.

127. Bagai A, Parsons K, Malone B, Fantino J, Paszat L, Rabeneck L. Workplace colorectal cancer–screening awareness programs: an adjunct to primary care practice? J Commun Health. 2007;32(3):157–67.

128. Allen JD, Stoddard AM, Mays J, Sorensen G. Promoting breast and cervical cancer screening at the workplace: results from the Woman to Woman Study. Am J Public Health. 2001;91(4):584.

129. Myers RE, Vernon SW, Tilley BC, Lu M, Watts BG. Intention to screen for colorectal cancer among white male employees. Prev Med. 1998;27(2):279–87.

130. Lazovich D, Parker DL, Brosseau LM, Milton FT, Dugan SK, Pan W, et al. Effectiveness of a worksite intervention to reduce an occupational exposure: the Minnesota wood dust study. Am J Public Health. 2002;92(9):1498–505.

131. Hogg-Johnson S, Robson L, Cole DC, Amick BC, Tompa E, Smith PM, et al. A randomised controlled study to evaluate the effectiveness of targeted occupational health and safety consultation or inspection in Ontario manufacturing workplaces. Occup Environ Med. 2012;69(12):890–900.

132. Eakin J. Work related determinants of health behavior. In: Gochman D, editor. Handbook of Health Behavior Research I: Personal and Social Determinants. New York: Plenum Press; 1997. pp. 337–57.

133. Sorensen G, Stoddard A, Hammond SK, Hebert JR, Avrunin JS, Ockene JK. Double jeopardy: workplace hazards and behavioral risks for craftspersons and laborers. Am J Health Promot. 1996;10(5):355–63.

134. Albertsen K, Hannerz H, Borg V, Burr H. Work environment and smoking cessation over a five-year period. Scand J Public Health. 2004;32(3):164–71.

135. Sorensen G, Stoddard A, Quintiliani L, Ebbeling C, Nagler E, Yang M, et al. Tobacco use cessation and weight management among motor freight workers: results of the gear up for health study. Cancer Cause Control. 2010;21(12):2113–22.

136. Sorensen G, McLellan D, Dennerlein JT, Pronk NP, Allen JD, Boden LI, et al. Integration of health protection and health promotion: rationale, indicators, and metrics. J Occup Environ Med. 2013;55(12):12–8.

137. Baron SL, Beard S, Davis LK, Delp L, Forst L, Kidd-Taylor A, et al. Promoting integrated approaches to reducing health inequities among low-income workers: applying a social ecological framework. Am J Ind Med. 2013.

138. Howard J, Hearl F. Occupational safety and health in the USA: now and the future. Ind Health. 2012;50(2):80–3.

139. Centers for Disease Control and Prevention. NIOSH Total Worker Health. 2013. Retrieved September 13, 2013 from http://www.cdc.gov/niosh/TWH/.

140. Campbell MK, Hudson MA, Resnicow K, Blakeney N, Paxton A, Baskin M. Church-based health promotion interventions: evidence and lessons learned. Annu Rev Public Health. 2007;28:213–34.

141. Kosmin BA, Keysar A. Religion in a free market: religious and non-religious Americans who/what/why/where. Ithaca, NY: Paramount Market Publishers; 2006.

142. Kosmin BA, Keysar A, Cragun R, Navarro-Rivera J. American nones: the profile of the no religion population, a report based on the American Religious Identification Survey 2008. 2009.

143. Drake BF, Shelton R, Gilligan T, Allen JD. A church-based intervention to promote informed decision-making for prostate cancer screening among African-American men. J Natl Med Assoc. 2010;102(3):164.

144. Resnicow K, Campbell M, Carr C, McCarty F, Wang T, Periasamy S, et al. Body and soul: a dietary intervention conducted through African-American churches. Am J Prev Med. 2004;27(2):97–105.

145. Fuemmeler BF, Mâsse LC, Yaroch AL, Resnicow K, Campbell MK, Carr C, et al. Psychosocial mediation of fruit and vegetable consumption in the body and soul effectiveness trial. Health Psychol. 2006;25(4):474.

146. Allen JD, Mars DR, Tom L, Apollon G, Hilaire D, Iralien G, et al. Health beliefs, attitudes and service utilization among Haitians. J Health Care Poor U. 2013;24(1):106–19.

147. Allen JD, Pérez JE, Pischke CR, Tom LS, Juarez A, Ospino H, et al. Dimensions of religiousness and cancer screening behaviors among church-going Latinas. J Relig Health. 2012:1–14.

148. Allicock M, Campbell MK, Valle CG, Carr C, Resnicow K, Gizlice Z. Evaluating the dissemination of Body & Soul, an evidence-based fruit and vegetable intake intervention: challenges for dissemination and implementation research. J Nutr Educ Behav. 2012;44(6):530–8.

149. Ferreira I, van der Horst K, Wendel-Vos W, Kremers S, van Lenthe FJ, Brug J. Environmental correlates of physical activity in youth—a review and update. Obes Rev. 2007;8(2):129–54.

150. van der Horst K, Oenema A, Ferreira I, Wendel-Vos W, Giskes K, van Lenthe F, et al. A systematic review of environmental correlates of obesity-related dietary behaviors in youth. Health Educ Res. 2007;22(2):203–26.

151. Birch LL, Fisher JO. Development of eating behaviors among children and adolescents. Pediatrics. 1998;101(3 Pt 2):539–49.

152. Lien N, Lytle LA, Klepp KI. Stability in consumption of fruit, vegetables, and sugary foods in a cohort from age 14 to age 21. Prev Med. 2001;33(3):217–26.

153. Ashcroft J, Semmler C, Carnell S, van Jaarsveld CH, Wardle J. Continuity and stability of eating behaviour traits in children. Eur J Clin Nutr. 2008;62(8):985–90.

154. Janz KF, Dawson JD, Mahoney LT. Tracking physical fitness and physical activity from childhood to adolescence: the muscatine study. Med Sci Sport Exer. 2000;32(7):1250–7.

155. Hill KG, Hawkins JD, Catalano RF, Abbott RD, Guo J. Family influences on the risk of daily smoking initiation. J Adolescent Health. 2005;37(3):202–10.

156. Turner L, Mermelstein R, Flay B. Individual and contextual influences on adolescent smoking. Ann NY Acad Sci. 2004;1021:175–97.

157. Golan M, Fainaru M, Weizman A. Role of behaviour modification in the treatment of childhood obesity with the parents as the exclusive agents of change. Int J Obes Relat Metab Disord. 1998;22(12):1217–24.

158. Golan M, Weizman A, Apter A, Fainaru M. Parents as the exclusive agents of change in the treatment of childhood obesity. Am J Clin Nutr. 1998;67(6):1130–5.

159. Small S, Huser M. Family-based prevention programs. In: Levesque R, editor. Encyclopedia of adolescence. New York: Springer; 2012.

160. Davison KK, Jurkowski JM, Li K, Kranz S, Lawson HA. A childhood obesity intervention developed by families for families: results from a pilot study. Int J Behav Nutr Phys Act. 2013;10:3.

161. Jurkowski J, Greenpope L, Lawson H, Bovenzi M, Quartimon R, Davison K. Engaging low-income parents in childhood obesity prevention from start to finish: a case study. J Commun Health. 2013;38:1–11.

162. Fetterman D, Wandersman A. Empowerment evaluation yesterday, today and tomorrow. Am J Eval. 2007;28(2):179–98.

163. Ackermann RT, Finch EA, Brizendine E, Zhou H, Marrero DG. Translating the diabetes prevention program into the community: the DEPLOY Pilot Study. Am J Prev Med. 2008;35(4):357–63.

164. Ackermann RT, Marrero DG. Adapting the Diabetes Prevention Program Lifestyle Intervention for Delivery in the Community the YMCA model. Diabetes Educator. 2007;33(1):69–78.

165. Battaglia TA, Bak SM, Heeren T, Chen CA, Kalish R, Tringale S, et al. Boston patient navigation research program: the impact of navigation on time to diagnostic resolution after abnormal cancer screening. Cancer Epidem Biom Prev. 2012;21(10):1645–54.

166. Freeman HP. Patient navigation: a community centered approach to reducing cancer mortality. J Cancer Educ. 2006;21(1 Suppl):S11.

167. Freund KM. Patient navigation: the promise to reduce health disparities. J Gen Intern Med. 2011;26(2):110–2.

168. Fisher EB, Boothroyd RI, Coufal MM, Baumann LC, Mbanya JC, Rotheram-Borus MJ, et al. Peer support for self-management of diabetes improved outcomes in international settings. Health Affair. 2012;31(1):130–9.

169. Boothroyd RI, Fisher EB. Peers for progress: promoting peer support for health around the world. Fam Pract. 2010;27(suppl 1):i62–i8.

170. Havas S, Heimendinger J, Damron D, Nicklas TA, Cowan A, Beresford SA, et al. 5 A Day for better health—nine community research projects to increase fruit and vegetable consumption. Public Health Rep. 1995;110(1):68.

171. Webber D, Balsam A, Oehlke B. The Massachusetts farmers' market coupon program for low income elders. Am J Health Promot. 1995;9(4):251–3.

172. Eyler AA, Brownson RC, Aytur SA, Cradock AL, Doescher M, Evenson KR, et al. Examination of trends and evidence-based elements in state physical education legislation: a content analysis. J School Health. 2010;80(7):326–32.

173. Land T, Rigotti NA, Levy DE, Paskowsky M, Warner D, Kwass J-A, et al. A longitudinal study of medicaid coverage for tobacco dependence treatments in Massachusetts and associated decreases in hospitalizations for cardiovascular disease. PLoS Medicine. 2010;7(12):e1000375.

174. Land T, Warner D, Paskowsky M, Cammaerts A, Wetherell L, Kaufmann R, et al. Medicaid coverage for tobacco dependence treatments in Massachusetts and associated decreases in smoking prevalence. PloS One. 2010;5(3):e9770.

175. Connolly G, Robbins H. Designing an effective statewide tobacco control program-Massachusetts. Cancer. 1998;83(S12A):2722–7.

176. Koh HK, Judge CM, Robbins H, Celebucki CC, Walker DK, Connolly GN. The first decade of the Massachusetts Tobacco Control Program. Public Health Rep. 2005;120(5):482.

177. Biener L, Harris JE, Hamilton W. Impact of the Massachusetts tobacco control programme: population based trend analysis. BMJ. 2000;321(7257):351–4.

178. Green ED, Guyer MS. Charting a course for genomic medicine from base pairs to bedside. Nature. 2011;470(7333):204–13.

179. Petronis A. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. Nature. 2010;465(7299):721–7.

180. McBride CM, Bryan AD, Bray MS, Swan GE, Green ED. Health behavior change: can genomics improve behavioral adherence? Am J Public Health. 2012;102(3):401–5.

181. Hay J, Baguer C, Li Y, Orlow I, Berwick M. Interpretation of melanoma risk feedback in first-degree relatives of melanoma patients. J Cancer Epidemiol. 2012;2012.

182. Altman DG. Sustaining interventions in community systems: on the relationship between researchers and communities. Health Psychology. 1995;14(6):526.

CHAPTER 11

EXPERIMENTAL PSYCHOSOCIAL INTERVENTIONS

Thomas A. Glass, Amii M. Kress, and Lisa F. Berkman

INTRODUCTION

Social epidemiology is often asked about the usefulness or policy relevance of its research product. Beyond the mission to document social inequalities in health, there is a need to translate findings into interventions that improve population health. Epidemiology has long been a pragmatic enterprise, driven by the need not only to understand how patterns of disease arise, but also to determine which pump handles to remove. However, translating what we learn in social epidemiology into interventions is a daunting task. The interventions that stem from social epidemiology often clash with powerful vested political, economic, and social interests. The focus on higher (mezzo- and macroscopic) levels of influence leads to the need for interventions at higher levels of organization. Such interventions can be more costly, more difficult to undertake and evaluate, and more politically caustic. On the other hand, such interventions have the advantage of actually identifying implementable policy and practice changes and, secondly, may be methodologically stronger than observational studies if they employ strong experimental designs. For these latter reasons, investigators are increasingly making the transition from observational studies to interventions designed to improve health and functioning. As we shall see, this transition has not always been smooth. The design, evaluation, and implementation of psychosocial interventions are an exceptionally challenging enterprise, and the field remains relatively new. Since the first edition of this book, a number of important intervention trials have emerged. However, such trials remain expensive and challenging to implement in a manner that remains true to the theories and findings that motivated them.

Social epidemiology exists within epidemiology, where experimental studies retain their exalted place as the king of evidence. As our causal language has evolved, the potential outcomes

framework has surpassed more classical ideas (e.g., Bradford Hill's causal viewpoints) (1, 2). This framework requires us to imagine a comparison between two possible worlds, one in which some experimental manipulation has taken place, and the other, identical in other respects, without that manipulation. Experimental logic has thus permeated our thinking well beyond the scope of studies that are explicitly evaluating interventions. It is not surprising then that social epidemiology turns to our most reliable and robust tool for evaluating interventions: the experiment.

However, the randomized clinical trial (RCT), recognized as the optimal experimental design, is often ill suited to the testing of interventions involving social processes and exposures that operate at higher levels of organization and that impact communities or entire societies. The RCT is optimized for the evaluation of individually targeted pharmacologic interventions that can be neatly separated from social context, can be administered in standard dosage, without contamination, and in which blinding can be achieved simply with a sugar pill. The challenges of adopting this design to the study of structural interventions, or manipulations of "fundamental causes" are numerous and well documented (3–5). However, if social epidemiology is correct in arguing that the important drivers of population health are larger social structural factors at higher levels of organization, then it is imperative that we develop convincing and rigorous approaches to evaluation. Otherwise, the field will study only interventions that are easiest and cleanest to fit into the RCT mold. This results in *lamp-post* bias, looking exclusively at factors that are easiest to study (where the light shines) despite knowledge that the important drivers of population health lie elsewhere (in the dark).

To date, experimental studies testing the effect of psychosocial interventions have produced mixed results. However, progress in the development of new interventions and new methodologies can now be seen. Psychosocial intervention studies are expensive, time-consuming, and complex undertakings that require careful planning and clear conceptualization. A well-executed experimental study of a psychosocial intervention provides powerful and compelling evidence for a causal link between social and behavioral factors and disease etiology. Further, experimental studies can help to identify the conditions under which candidate social factors are alterable, thereby providing clues about pathways through which social factors operate. Intervention studies also provide an evidentiary basis for the implementation of policies and programs that address what Rose called "mass influences" that are key contextual determinants of the sentinel health behaviors that underlie population health (6). Decades of experience have taught us that changing behavior is difficult and does not always lead to desired and anticipated improvements in health. With these challenges and promises in mind, the goals of this chapter are to:

1. delineate the boundaries and characteristics of experimental psychosocial interventions;
2. selectively highlight previous studies;
3. propose a series of theoretical tools to guide the next generation of intervention studies;
4. summarize the major methodological and conceptual pitfalls and some tentative ameliorative strategies; and
5. suggest areas of future research.

In the course of addressing these goals, the chapter will be guided by five propositions, both conceptual, and methodological. These propositions summarize important lessons learned to date,

and are intended to guide researchers in the development of future experimental studies of psychosocial intervention. These guiding propositions are:

1. **Explicate the theoretical underpinnings of the intervention.** Detailed attention to issues of theory will guide the selection of variables, the choice of intervention strategies, and the design. The need for theory exists on several distinct levels. Previous research has suffered from inattention to "upstream" factors through which the larger social context affects individual behavioral and psychological factors.

2. **Target a strategic psychosocial mechanism shown to be related to the health outcome.** Successful intervention designs focus on a circumscribed set of specific and strategically chosen behavioral processes with a clear link to the health outcome of interest. If observational research has not demonstrated a link between the intended target and the outcome of interest, an intervention study is most likely premature.

3. **Choose a well-accepted and psychometrically sound measure of health or functioning as an outcome.** Psychosocial outcomes, such as coping, adjustment, or well-being are important, however, far less compelling than are "hard" outcomes in which health or functioning are directly measured. Self-reported psychosocial outcomes are also more prone to same-source bias, given that many psychosocial interventions cannot be blinded. Previous research suggests that behavior change does not necessarily lead to expected changes in health. Therefore, while behavior change is a suitable outcome on its own, health outcomes remain the gold standard.

4. **Calibrate the intervention to the lifecourse.** Many interventions have failed because they gave the right intervention at the wrong developmental period or delivered an intervention that was underpowered given the developmental trajectory of the psychosocial process. A key lesson of the past decade has been the importance of etiologic periods with regard to both the exposures and outcomes of interest.

5. **Strive for the strongest possible experimental design.** Although the methodological challenges of pharmacological trials and psychosocial intervention trials are largely analogous, social epidemiologists can expect exceptional methodological scrutiny. For this reason, investigators should select the strongest possible designs. The most compelling evidence for novel therapies will come from the most rigorous study designs. Sample sizes should be carefully considered, and when possible, randomized double-blind trials should be conducted. The unit of randomization can vary from individuals to more area-based units such as floors of a building, schools, worksites, and communities. Group-level randomization procedures including multilevel clusters (classrooms with schools, worksites within companies, counties within states) may be more appropriate to social interventions than single-level designs.

WHAT IS PSYCHOSOCIAL INTERVENTION?

The term *psychosocial intervention* is used widely in many fields including nursing, psychology, psychiatry, social work, sociology, and behavioral science. In its simplest meaning, it refers to a systematic attempt to modify a psychosocial process. By psychosocial process, we refer broadly to a family of social and/or psychological factors that are known to impact health directly

(e.g., social isolation, job strain, social capital, inequality, discrimination) or indirectly, through their influence of health-related behaviors (e.g., self-efficacy, peer pressure, cultural norms, risky or health-promoting behaviors such as care-seeking, tobacco or alcohol consumption). Psychosocial factors are intermediate steps in multilevel causal chains that link macro-structural factors to behaviors and health states. They are key mechanisms through which fundamental causes get "under the skin." Interventions to alter psychosocial processes can occur at the level of the individual, the family, the social network, the workplace, community, or at the population level. Defined in this way, changes in public policy designed to modify behavior, such as increased taxation of tobacco, constitute psychosocial intervention. As the field has evolved over the last decades, the meaning of psychosocial has expanded beyond psychology to include concepts and principles from behavioral economics (see Chapter 13) to biology (see Chapter 14) to broader aspects of social and public policy (see Chapters 4, 6, and 15). This tracks with a broader trend toward an increasingly transdisciplinary understanding of the social conditions that enable and constrain how individuals make behavioral choices based on context.

In this chapter, we examine psychosocial intervention as a core function of the mission of social epidemiology. *The focus is on experimental interventions aimed at changing some psychosocial process for the explicit purpose of modifying physical health or functioning.* This includes the primary prevention of disease onset, recovery from illness, secondary prevention of disease, as well as modification of the course of disease. We will not cover psychiatric and psychological interventions that explicitly target mental health outcomes. We also neglect interventions designed to modify physiologic mechanisms without explicit linkage to the health consequences of those changes (e.g., meditation and relaxation). We also will not review studies of interventions that target behavior change alone, without explicit attention to the resulting health impact (Chapter 10 covers many excellent approaches to health behavioral change frameworks). Finally, while many psychosocial interventions involve health education, we do not review the extensive body of research on interventions to modify health knowledge and attitudes alone (for excellent reviews of these studies see 7, 8, 9). Intervention studies not included in this review are important in their own right, and no invidious comparison is intended. Many aim to modify health states or to alter disease risk in individuals or populations, but extensive reviews exist outside of epidemiology in various branches of behavioral science and are beyond the scope of the current chapter (see Chapter 10). Instead, the focus will be on seven types of psychosocial interventions described below.

A TYPOLOGY OF PSYCHOSOCIAL INTERVENTIONS

There are several potential bases of organizing a typology of psychosocial interventions. One approach would be to emphasize the factors targeted for manipulation, or the desired outcome. Another approach would be to categorize according to the intended target population or disease stage (e.g., primary prevention, illness recovery, etc.). For purposes of this chapter, seven types of intervention studies will be reviewed: *(1)* behavioral change interventions, *(2)* social support interventions, *(3)* disease management interventions, *(4)* distress mitigation interventions,

(5) control/efficacy enhancement interventions, *(6)* collective efficacy in communities, and *(7)* organizational change interventions aimed at organizations and workplaces. This typology emphasizes the psychosocial mechanism targeted as well as whether the intervention is focused on prevention, or alteration of the course of disease. The separations between types are often blurred. This categorization is for heuristic purposes only as a way of organizing studies to date. A selective sample of influential studies in each category is presented in Table 11.1. This is not an exhaustive list; many studies that lacked rigorous evaluation or that produced negative or inconclusive findings have been omitted. The emphasis is on note-worthy illustrations of each intervention type rather than on a comprehensive review.

BEHAVIORAL CHANGE INTERVENTIONS

The largest cluster of studies reviewed had as their goal the modification of specific health-related behaviors found to be risk factors for disease onset or recurrence. Most of these have been primary or secondary prevention efforts aimed at cardiovascular disease (for reviews see 10–20). Overall, the performance of population-based primary prevention trials designed to alter "lifestyle" factors has been mixed. Little or no benefit in long-term follow-up has been observed in the Goteborg Primary Prevention Trial (21), the Minnesota Heart Health Program (MHHP) (22, 23), and the Pawtucket Heart Health Program (24). Disappointingly small changes in health behaviors have been observed in the Stanford Five-City Multi-Factor Risk Reduction Project (FCP) (25, 26), and the WHO European Collaborative Group Trial (27). Other trials aimed at high-risk individuals, including the Multiple Risk Factor Intervention Trial (MRFIT), observed small changes in health behaviors that did not translate into anticipated reductions in rates of morbidity or mortality (28).

Numerous commentaries have been offered to explain why these large, high-visibility studies fell short of expectation, such as the exceptional summary editorial by Susser (29). Susser notes that many community-based trials have failed to overcome large-scale social movements resulting in changes in the control subjects that, in turn, impacted the power of the trial to detect real improvements. Or as Susser put it, the trials were "[outrun] by the pace of social change" (p. 157).

In most of these trials, the health behaviors that constitute risk factors for disease (diet, smoking, exercise) have been treated as discrete, voluntary, and individually modifiable "lifestyle" choices, detached from the social context in which behaviors arise (30). For this reason, many of these trials have been criticized for ignoring "upstream" social factors antecedent to behaviors at an individual level. This tendency to ignore the contextual basis of behaviors is reflected in the theoretical foundations on which many of the primary prevention trials have been formulated. Often, no conceptual basis for the intervention is articulated. Among those studies that have made their theoretical models explicit, most appear to have been influenced by some variant of social learning theory (see 31). Social learning theory, to the extent that it emphasizes self-efficacy beliefs as properties of individuals, tends to shift the focus away from upstream factors related to the social context, toward more individually based models. This has contributed to what Rockhill terms the privatization of risk (32), and to the notion that health behaviors are discrete, atomized, and can be changed without regard to the larger social context.

A number of other trials did show evidence of risk factor reduction as well as declines in subsequent coronary heart disease (CHD) morbidity. In the Oslo trial, 5-year CHD incidence was

TABLE 11.1: Selected list of psychosocial intervention studies by type

Principal Author, year. Title.	Study Design	Intervention	Main Results	General Comments
1. Behavior Change Interventions				
Anonymous, 1982 "Multiple Risk Factor Intervention Trial (MRFIT)" (310)	Randomized primary prevention trial of 12,866 high-risk middle-aged men recruited at 22 sites. Average duration of follow-up was 7 years. 361,662 men were screened, and 3.5% were enrolled. Seven percent qualified on the basis of risk screening	Special intervention included stepped-care treatment for hypertension, counseling for cigarette smoking, dietary advice for lowering blood cholesterol	Risk factors decreased more for treatment group than control group, but differences were modest and may not have been sustained. Overall CHD mortality differences were nonsignificant. Overall mortality rate 2% higher in treatment group	Trial criticized for not addressing social and environmental factors and for exclusive emphasis on high-risk individuals. Variation in risk reduction among high-risk individuals was high. Secular trends in risk factor reduction reduced statistical power from 90% to 60%. Demonstrates limitations of individually targeted "highrisk" approach
Anonymous, 1995. "Community Intervention Trial for Smoking Cessation (COMMIT)" (311, 312)	Heavy smokers, n = 10,019; light-to-moderate smokers, n = 10,328	Community-level·multichannel 4-year intervention designed to increase smoking cessation among 11 matched community pairs (10 in US, 1 in Canada)	Mean quit rate for heavy smokers who received intervention was 0.180. Mean quit rate for comparison group was 0.187; nonsignificant difference. Significant differences for light to-moderate smokers in intervention (0.306) and comparison (0.275) communities	Women less likely than men to be heavy smokers but twice as likely to feel pressure to quit

(*continued*)

TABLE 11.1: Continued

Principal Author, year. Title.	Study Design	Intervention	Main Results	General Comments
Carleton, 1995. "Pawtucket Heart Health Program" (24)	Randomized community-based intervention	Community-wide education program applied to intervention city on three levels: risk factors, behavior change, and community activation	Statistically significant city decrease in projected cardiovascular disease rate for Pawtucket (16%) during peak intervention time. This decrease was not maintained after postintervention period	No long-term discernible effects on mortality or risk factors. Attributable to mass media messages to which both towns may have been exposed
Farquhar, 1990. "Stanford Five-City Multi-Factor Risk Reduction Project (FCP)" (25, 313)	A 14-year trial of community wide CVD risk reduction. Nonrandomized community level intervention in northern CA. Epidemiologic surveillance and measurement of health-related behaviors over 14-year period (n = 122,800 in treatment cities; n = 197,500 in control cities)	Community-wide organization and health education including media and "personal influence" in tx communities lasting 5 years. Targets included (1) lower plasma cholesterol through diet change, (2) reduced blood pressure, (3) weight control, and (4) increased physical activity. Spanish-language program also implemented	Net decrease in mean cholesterol level (2%), and mean blood pressure (4%) in both cohort and independent samples. These risk factor changes resulted in composite total mortality risk scores that were 15% lower in the intervention communities and CHD risk scores that were 16% lower	Positive intervention effect on risk factors. Results may not be generalizable since samples were not randomly chosen or assigned. Cities are intervention units. Results of independent samples differ from cohort samples. This may be due to less exposure to education (e.g., recent immigrants)

Friedman, 1984. "Recurrent Coronary Prevention Project (RCCP)" (39–42, 314)	Post-MI patients treated for 4.5 years, after which subjects were followed for additional 4 years. At beginning of trial control group n = 270, experimental group n = 592, comparison group n = 151	Control group received group cardiac counseling. Experimental group received group cardiac counseling plus type A behavioral counseling. Comparison group did not receive either type of counseling	At end of 4.5 years 35.1% of treatment group showed "markedly reduced" type A behavior. Control group decreased by 9.8%. Cumulative recurrence rate of MI of treatment group was 12.9% (compared with 21.2% cumulative recurrence rate of control group or 28.2% cumulative recurrence rate of comparison group)	
Kelly, 1991. "HIV Risk Behavior Reduction Following Intervention with Key Opinion Leaders of Population: An Experimental Analysis" (315)	One intervention city and two comparison cities (pop = 50,000–75,000 residents). Surveys of male patrons of clubs completed at baseline, 3 months (intervention n = 328; comparison n = 331), and 6 months after training period (intervention n = 278; comparison n = 330)	Key opinion leaders chosen by bartenders at clubs and were trained in HIV risk reduction behavior, strategies, and role playing	Intervention city men who engaged in unprotected anal intercourse decreased (–25% from baseline), a reduction of unprotected receptive anal intercourse (–30%), an increase in condom use (16%) during anal intercourse, and a decrease in percentage of men with more than one sexual partner (–18%)	Generalizable to population but not to those who do not frequent clubs, nonwhites, and teens. Bias-self-reported results.

(*continued*)

TABLE 11.1: Continued

Principal Author, year. Title.	Study Design	Intervention	Main Results	General Comments
Hjermann, 1983. "The Oslo Trial." (33)	Randomized trial (5 years) of healthy, normotensive, but coronary-high-risk men (intervention n = 604, control n = 628)	Subjects were advised individually for 10–15 minutes by a doctor to stop smoking and to lower their blood lipids by dietary changes	Associated with 13% lower mean serum cholesterol, 20% decrease in mean fasting serum triglycerides, 45% reduction in tobacco consumption, and 47% lower incidence of MI (fatal and nonfatal) and sudden death.	
Levenkron, 1983. "Modifying the Type A Coronary-Prone Behavior Pattern" (316)	Male volunteers (total n = 38) between the ages of 25 and 50 received treatment. Comprehensive behavior therapy (CBT), n =12; group support (GS), n =13; brief information (BI), n =13	CBT group received training in self-control and relaxation. GS group encouraged self-awareness of TABP, specification of Type-A behaviors, and "inducing change through non-specific support and exhortation of both therapist and group members."	CBT and GS groups showed decreases in TABP components (e.g., Jenkins Activity Survey, Hard driving, Job involvement factor, Framingham, and Type A Scale.) Trend in negative mean changes in plasma-free fatty acids for CBT and GS groups. CBT group exhibited significant decrease in triglycerides	Results may not be generalizable to population-subjects mainly from one corporation and were healthy, highly educated, and nondistressed males. Unexpected result: serum cholesterol increased across all groups

Lindstrom, 2003. **"Finish Diabetes Prevention Study (DPS)" (317–320)**	Randomized controlled trial of overweight middle-aged adults with impaired glucose tolerance (n = 522). Participants were randomized to intensive lifestyle intervention (n = 265) or control (n = 257)	The intervention lasted from less than one year up to six years and included detailed and individualized counseling to achieve lifestyle goals. Intervention included seven personal counseling sessions with the study nutritionist during the first year and every three months thereafter. Participants were also advised to increase their level of physical activity, and were offered free of charge, supervised individually tailored circuit-type moderate intensity resistance training sessions. Participants in the control group were given general information about lifestyle and diabetes risk either individually or in one group session and were provided with some printed materials. Control group counseling was not individualized	The study was prematurely terminated by an independent end point committee because the incidence of diabetes in the intervention group was significantly lower than in the control. During follow-up the incidence of type 2 diabetes was 4.3 and 7.4 person-years in the intervention and control group, respectively, indicating a 43% reduction in risk associated with the intervention. Associated with weight reductions after 1 and 3 years, weight reductions were 4.5 and 3.5 kg in the intervention group and 1.0 and 0.9 kg in the control group, respectively. Measures of glycemia and lipemia improved more in the intervention group	

(*continued*)

TABLE 11.1: Continued

Principal Author, year. Title.	Study Design	Intervention	Main Results	General Comments
Luepker, 1994. "Minnesota Heart Health Program (MHHP)" (23, 321)	13-year community-wide research and demonstration project	Three pairs of marched intervention and comparison communities received 5-year health education program. Program designed to improve health behaviors, lower blood cholesterol and blood pressure, and reduce cardiovascular disease morbidity and mortality	No discernible differences between intervention communities and comparison communities	Results attributed to secular trends of increasing health promotion and declining risk factors
Ornish, 1983. "Lifestyle Heart Trial" (12, 43, 44, 322)	Randomized controlled clinical trial to rest short-term effects of lifestyle on coronary heart disease	Intervention to reach lifestyle changes (low-fat vegetarian diet, smoking cessation, stress management, moderate exercise). 2 x weekly group discussions provided social support to facilitate lifestyle changes. Discussions led by clinical psychologist; promoted adherence to program, communication skills, & expression of feelings	82% of experimental group had an average change toward regression of coronary artery lesion diameters. Greater changes were found in more severely stenosed lesions	Lifestyle changes may be effective without use of lipid-lowering drugs. This study also had social support features

Pahkala, 2013. **"STRIP Study (Special Turku Coronary Risk Factor Intervention Project)"** (323)	Prospective randomized study designed to examine influence of saturated fat-oriented counseling on atherosclerosis in children. Children were recruited at their 5-month visit at the Turku City well-baby clinic (intervention n = 540; control n = 522). The intervention was designed to stop at 20 years of age.	Dietary counseling of intervention families was given by a nutritionist at every visit (3- to 12-month intervals). Children and family were involved in counseling. During the first year, the intervention was aimed at the parents and child-level interventions started at age 7	Favorable effect on serum cholesterol values and endothelial function. At age 15, children in the intervention group had better cardiovascular health	
Patterson, 2006. **"Functional Adaptation Skills Training (FAST)"** (324)	Randomized controlled trial of 240 patients with a DSM-IV-based chart diagnosis of schizophrenia or schizoaffective disorder. Trial was 24 weeks. The intervention (n = 124) was compared with a time-equivalent attention-control group (n = 116)	Intervention was a behavioral group intervention based on social cognitive theory and the Social and Independent Living Skills Program. It targeted six areas of everyday functioning: medication management, social skills communication skills, organization and planning, transportation, and financial management. Intervention participants met in a group format for 120 minute once per week over the course of 24 weeks	Significant improvements in everyday living skills and social skills but not medication management	

(*continued*)

TABLE 11.1: Continued

Principal Author, year. Title.	Study Design	Intervention	Main Results	General Comments
Puska, 1989. "The North Karelia Project" (325) See also (326–331)	Comprehensive communitybased health intervention launched in 1972. Results compiled through survey of representative samples of individuals in 3 communities-North Karelia (intervention) and Kuopio County and south west Finland at 5-year intervals	Program targeted reductions in risk factors (smoking, serum cholesterol, and blood pressure). Comprehensive health education program to promote healthy life styles, taught practical skills, provided social support for change, and arranged environmental modifications. Unit of intervention was community.	After 10 years, reductions in smoking (28%), hypertension (3%), and serum cholesterol levels (3%) were faster among men in intervention community than in comparison community or rest of Finland. Significant improvements in women observed only for blood pressure. Between 1974 and 1979, CHD mortality declined twice as fast in N. Karelia (22%) compared with reference community (12%) or rest of Finland (11%) ($p < 0.05$)	First large-scale primary prevention demonstration (on which many other studies were based). Study launched in 1972 after public outcry over statistics showing that Finnish men had the highest rates of CHD mortality and risk factor prevalence in Europe. Study demonstrates importance of community approach as opposed to "high risk" strategy (see Rose). Among the few studies that attempted environmental modifications along with traditional health education model. May have been effective because it predated onset of significant secular trend toward reductions in risk factors

significantly reduced (by 47%) in the intervention group compared with the control group (33). The first such community-based CHD prevention project was the North Karelia Project, which involved a "comprehensive community organization for change" including individual behavior change interventions complemented by social supports and environmental modifications (31, 34–37). This study was widely influential, in part because it demonstrated both reductions in risk factors (smoking, blood pressure, and serum cholesterol) and reductions in both morbidity and mortality. The success of North Karelia may have resulted in part from the accident of timing: it came before large-scale secular trends toward increased exercise, dietary change, and smoking cessation had fully begun (38). Interestingly, in the aftermath of successful results in North Karelia, implementation of heart disease prevention efforts in the rest of Finland was slow to take root (38). One important feature of the North Karelia Project was that it included broader structural changes at higher levels of organizations, such as changes in incentives for milk and sausage producers to lower fat content, or to switch to fruit production. North Karelia was a multilevel intervention that coupled behavioral interventions targeting individuals with support programs and health education targeted at community organizations. These structural and environmental features were not generally copied in the subsequent (and less effective) primary prevention projects undertaken in the United States and elsewhere.

In contrast to large-scale community interventions that targeted a wide range of health behaviors all at once, a series of more focused, theoretically grounded studies has sought to target specific psychosocial processes. For example, several trials focused on altering type-A behavior (TAB). The most important intervention study in this area has been the Recurrent Coronary Prevention Project (RCPP) (39–41). In that study, 1,035 male and female subjects age 64 years or younger who had suffered their first or last documented acute myocardial infarction (MI) were enrolled in a 5-year study to determine the prevalence of TAB, the extent to which TAB itself is modifiable, and whether a program designed to alter TAB would result in lower rates of fatal and nonfatal coronary recurrence. After 1 year, results indicated that the prevalence of TAB was quite high (98%) and that rates of cardiovascular death and reinfarction were lower among subjects who received cardiologic and behavioral counseling compared with usual care controls (41). After 4.5 years of follow-up, 35 percent of the treatment group showed "markedly reduced TAB" compared with a decrease of 10% in the control group (42). Cumulative recurrence of MI was significantly lower (13% vs. 21%). The RCPP succeeded in showing that TAB could be modified and that the resulting change translated into reduced rates of coronary events. It was also among the only studies that showed that the benefits of the intervention were persistent over at least 5.5 years (39).

Another important study involving a systematic attempt to alter behavior was the Lifestyle Heart Trial conducted by Dean Ornish and colleagues (43–45). In this randomized trial, 28 men were given a short-term comprehensive lifestyle intervention involving low-fat vegetarian diet, smoking cessation, stress management training, moderate exercise, and support group discussions led by a psychologist. The results showed that 82 percent of the experimental group showed evidence of regression of coronary artery lesions 1 year after intervention. Greater improvements were found in more severely stenosed lesions. Long-term follow-up has demonstrated reduced coronary plaque burden and improved risk factor profiles 1 and 5 years after intervention (12, 43). The Lifestyle Heart Trial is notable as one of the first trials to demonstrate the benefits of a complex, multimodal behavioral intervention using a "hard" physiologic outcome. The success and

visibility of Ornish's work have led to the widespread adoption of cardiac rehabilitation programs modeled on this study.

SOCIAL SUPPORT INTERVENTIONS

Social networks and support have been shown to be associated with mortality, with morbidity, with recovery, and with disease course in numerous previous studies (see Chapter 7 in this volume). Interventions designed to bolster support or to provide specialized types of support are a natural extension of these observational studies. While the exact mechanisms that underlie the association between social support and health are not known, the evidence is sufficient to warrant the development of innovative intervention strategies that are designed to impact particular pathways. Support interventions have been conducted primarily at the individual level (for a review, see 46); however, several noteworthy programs have attempted to enhance the supportiveness of relationships in worksites (47), in families (48), and in caregiving networks (49). Typically, social support interventions have been conducted in populations previously afflicted with a major illness such as heart disease, stroke, cancer, or HIV. Additional examples include treatment approaches for addictions (50, 51) and schizophrenia (52–54).

At least five modalities of support interventions can be distinguished: *(1)* professionally led support groups, *(2)* mutual support groups, *(3)* multifamily support groups, *(4)* support mobilization interventions, and *(5)* support substitution interventions. In various ways, and from various theoretical orientations, each attempts to bolster social support resources either by potentiating naturally occurring support systems, or through what Gottlieb has termed "grafted support."[1] More extensive discussions of the conduct of support interventions can be found in overviews by Biegel (56) and Gottlieb (55, 57).

In general, the literature regarding support interventions has produced mixed results. This is the result, in part, of methodological shortcomings including small sample sizes and weak designs. In a thoughtful review of support interventions in patients with rheumatoid arthritis, Lanza and Revenson (58) argue that these failures may be due to the lack of firm theoretical grounding. Because support is the product of relationships that develop and change slowly, the benefits of support interventions may be missed in short-term intervention studies. The Enhancing Recovery in Coronary Heart Disease (ENRICHD) Patients Study tested the effects of a psychosocial intervention, designed to increase social support and alleviate depression, on all-cause mortality and nonfatal infarction in patients with recently diagnosed acute MI who were depressed and/or had low social support (59). This study is discussed in greater detail in the distress mitigation section below. Network mobilization must coordinate with the crisis phase of the illness or risk missing the window of opportunity for maximal efficacy if we assume that is the time period where people are most likely to change. This, however, has proved to be challenging in execution. At least one network intervention effort failed because the intervention started too late (60).

1 Gottlieb refers to "grafted support" as support opportunities that are created by the intervention and are presumed to become efficacious sources of support over time.55. Gottlieb BH, editor. Marshalling Social Support. Beverly Hills, CA: Sage; 1988.

The Families in Recovery from Stroke Trial (FIRST) tested family-systems-based psychosocial intervention (PSI) designed to enhance recovery in stroke patients. The study randomized 291 stroke survivors to PSI or usual care within 28 days of index hospital admission (61). The intervention consisted of up to 16 in-home sessions, conducted by a psychologist or social worker trained in family-systems and cognitive behavioral therapy. Sessions included the stroke survivor's social network and were designed to increase self-efficacy, problem solving, social support, and cohesion. The intervention was not associated with improved functional recovery from stroke; however, preplanned subgroup analyses suggested that the intervention was more effective in subjects who were healthier at baseline (62). An RCT of family support for stroke survivors (n = 323) and their caregivers (n = 267) in the UK also concluded that the intervention was not associated with improvements in disability or handicap (63). Reviews suggest that psychosocial interventions have beneficial impacts on psychological outcomes, particularly in stroke patients (64), but the benefits of psychosocial intervention on health and functional outcomes in stroke remains unclear.

Many interventions that do not identify themselves as support interventions contain significant support components that arise indirectly. For example, the Lifestyle Heart Trial (45) of "lifestyle" change in men with heart disease is classified as a behavioral change intervention, yet the support group organized as a vehicle to facilitate the behavioral interventions also provides an avenue for social support. A recent large randomized study of peer counseling and support in Malawi showed significant reductions in infant mortality after 3 years of follow-up (65). Separating the impact of group support from didactic training is difficult in studies of this type. It is impossible to capture the impact of social support or to separate its influence from the intended impact of other aspects of the intervention. In another example, in an intervention to teach arthritis patients self-care behaviors in a group setting, participants cited the "feeling of knowing that everyone cared" as a benefit of the intervention (66). The group leader also noted that participants "appeared to benefit psychologically from the emotional support given in the small-group setting" (66: p. 81). Many educational or behavior change interventions include support components that are difficult to measure (67–70).

One intervention trial that has received extensive attention by Spiegel and colleagues (71) randomized 86 patients with metastatic breast cancer to either a control group, or to a 1-year intervention consisting of weekly supportive group therapy with self-hypnosis for pain. At 10-year follow-up, women in the treatment group survived twice as long as those in the control group, even after controlling for stage at diagnosis, treatment differences, and several other factors (mean of 36.6 months vs. 18.9 months, respectively). Although the treatment consisted of both support and disease management features, this study provides some of the most powerful evidence to date that a support group model is associated with longer survival, even in terminally ill patients. The original study led to a larger multicenter replication (72), but after examining 14-year follow-up data, longer survival was not observed among women randomized to the intervention group (73). However, in that study, women with estrogen receptor-negative tumors did show longer survival. The authors theorize that advances in treatments for estrogen receptor-positive tumors may have diminished benefits of psychosocial intervention. More recent evidence has been mixed. For example, a randomized trial of group psychotherapy in 227 women with early breast cancer found lower rates of recurrence and longer survival over 11 years (74). A similar model has been shown to be associated with improvements in CD4 cell counts and viral load in patients with

HIV (75). However, several recent reviews also show that about equal numbers of studies have failed to find either survival or psychosocial benefits in cancer patients (76, 77). In some studies, outcomes appear to depend on initial severity and psychosocial vulnerability. For example, in a supportive-expressive group therapy intervention in metastatic breast cancer (78), benefits were seen only in women with the most baseline distress; women with little distress at baseline may actually have had worse psychosocial outcomes in the intervention group, suggesting the possibility that group support can have negative consequences for some patients.

DISEASE MANAGEMENT INTERVENTIONS

A third cluster of interventions target psychosocial aspects of the post-onset phase of illness in an effort to enhance the patient's ability to cope with the disease or to prevent symptom recurrence. Systematic reviews of this literature are available in cancer (79–92), diabetes (93–99), and chronic disease generally (100–102). The emphasis in these studies is on providing specific coping strategies designed to address particular problems encountered in the course of the disease in question. One group of studies has focused on improving adherence to medical treatments. In one important study by Richardson and colleagues (103), 94 newly diagnosed hematologic cancer patients were randomized to an intervention designed to improve compliance with chemotherapy or to a usual care control group. A combination of education, home-visits, and behavioral intervention was associated with longer survival even after controlling for a variety of factors including compliance. Lifestyle and disease management interventions have been shown to be as effective as pharmacological approaches in modifying cardiovascular disease risk in diabetes patients in the Diabetes Prevention Program (DPP) (104, 105).

In another nonrandomized study, Fawzy and colleagues evaluated a disease management program in patients with early diagnosis of malignant melanoma with good prognosis (106). The intervention included health education, problem-solving skills, stress-management techniques, and psychological support. Results showed improvement in active coping skills, and in significantly lower depression, fatigue, confusion, and total mood disturbance, as well as higher vigor in the intervention group. Also important was the finding that the intervention was associated with beneficial changes in one aspect of immune system performance: the NK lymphoid cell system (107). Long-term follow-up showed longer survival (108). Studies of HIV-positive men have failed to find improvement (or slower rates of deterioration) in immune parameters after psychosocial intervention (109). This may have been the result of underpowered testing ($n = 39$).

Another group of studies has focused on improving self-management skills for patients suffering from chronic illness. An illustrative example comes from Lorig and colleagues, who evaluated the Arthritis Self-Management Program developed at the Stanford Arthritis Center (110). This program included education, self-help groups, and home practice of self-management skills led by a trained lay educator. Family involvement was optional. The intervention was associated with increased knowledge, improved self-care behaviors, and decreased pain. A follow-up study showed that these benefits were long-lasting (111). Similar interventions have been launched in cardiac disease (112), stroke (113), caregiving in dementia (114–119), and arthritis (120).

DISTRESS MITIGATION INTERVENTIONS

A fourth type of psychosocial intervention that has shown promise aims to reduce or mitigate the distress associated with either onset or treatment consequences of serious illness. These studies have employed a wide variety of techniques for reducing distress from cognitive behavioral therapy, relaxation, and education, to careful screening and multimodal interventions that are tailored to particular patient needs. The primary target of change is the experience of stress/distress. Some focus more specifically on depression and/or depressive symptoms. Reviews and meta-analyses of these interventions in cardiovascular disease have been published by Desseldorp et al. (121) and Linden (122), showing robust effects on mortality.

The most important early study was the Ischemic Heart Disease Life Stress Monitoring Program (123, 124), in which 461 male patients recovering from myocardial infarction were randomized to usual care or to a stress-monitoring intervention. The hypothesis was that a coordinated program of screening and multimodal intervention would alter the risk of disease recurrence and death. Patients were interviewed over the phone to screen for signs of distress. Home-based individually tailored combinations of education, support, and referrals were conducted in men who scored high on this screen. One-year results showed significantly reduced distress symptoms in the treatment group. Controls were about twice as likely to die of cardiac causes compared with the intervention group (123). Mortality differences persisted after 7 years of follow-up, and appeared to be due primarily to sudden death of cardiac origin (124). Subsequently, Frasure-Smith's group published results of a large-scale extension of the earlier intervention, the Montreal Heart Attack Readjustment Trial (M-HEART), to reduce life-stress after MI in a cohort of 1,376 men and women, showing no benefits (although a significant increase in cardiac and all-cause mortality was observed in preplanned analyses among women in the intervention group) (125). Nevertheless, this study's influence has continued, due in part to pioneering innovations in two areas: *(1)* the concept of individually tailored interventions, and *(2)* the importance of risk screening. The risk-screening methodology has also been extended to application among women newly diagnosed with breast cancer (126).

The most important and widely discussed study in this area is the Enhancing Recovery in Coronary Heart Disease (ENRICHD) trial, which was the first multicenter psychosocial intervention trial sponsored by the National Heart, Lung and Blood Institute (NHLBI) (127). The ENRICHD project was among the largest and most ambitious psychosocial intervention trials ever undertaken, and a major step forward in the evolution of the field. ENRICHD was an RCT aimed at improving social support and reducing depression in post-MI patients (59, 128, 129). The primary outcome was the reduction of reinfarction and all-cause mortality. ENRICHD enrolled post-MI patients from more than 80 hospitals and 8 clinical centers across the United States (n = 2,481) who were either depressed and/or had low perceived social support. ENRICHD was based on evidence from a large number of longitudinal observational studies indicating that both depression and social support were related to survival post MI (130). Patients were randomized to an intervention arm involving cognitive behavioral therapy and enhanced social support or to usual care. Results from related clinical trials were inconsistent (42, 123, 131). As discussed, the results from M-HEART, a study (125) published just before the launch of ENRICHD, showed null results but suggested a trend that women in the intervention group fared worse than women in the usual care group ($p = 0.064$).

With a follow-up of 3.4 years using an intent-to-treat analysis, there were no differences between the intervention and control groups in ENRICHD ($p = 0.89$) (59). In fact, the survival curves completely overlap when looking at the primary endpoint of reinfarction or all-cause mortality. The ENRICHD study did find reductions in the mediating risk factors (depression and low social support), suggesting the intervention was changing the psychosocial processes on which it was designed to intervene (59). The magnitude and long-term differences were smaller than expected, however, causing concerns about the intervention's effectiveness. At the conclusion of the main intervention (at 6 months), however, significant differences between the intervention and usual care (UC) groups in social support and depressive symptoms were observed. Importantly, there were heterogeneous treatment effects: white men benefited from the intervention more than white women or black men or women. These results are discussed further in the section on subject recruitment and subgroup effects.

SENSE OF CONTROL INTERVENTIONS

Another group of studies describe interventions that seek to modify the individual's sense of control over events (for an excellent early review, see 132). What makes these studies of particular interest is that they often target higher-level environmental factors to induce a sense of mastery and control. These studies grew, in part, out of efforts to study how the hospital environment could be modified to better prepare patients for surgery (133). In a classic study by Klein et al. (134), environmental modifications designed to allow patients to feel more in control were put in place in a coronary care unit. Patients in the control-enhancing intervention evidenced fewer cardiovascular complications as well as lower catecholamine excretion. These studies, while smaller in number, are included in our typology because they are explicitly theory-driven, target a particular psychosocial mechanism linked to health, and (some) have used health measures as outcomes.

An early example of this type of study was conducted by Rodin and Langer (135) to encourage elderly nursing home residents to make a greater number of choices and to feel more control and responsibility for their own lives. The goal of the intervention was to slow declines in health and cognitive function often observed in institutionalized elderly patients by modifying residents' sense of control. Building on the tradition of environmental psychology, this study changed the physical and social environment in order to alter the targeted psychosocial factor. Residents in the intervention group were given a speech that emphasized that residents were responsible for themselves. The comparison group was told that they would be taken care of by the staff. The treatment group was given houseplants and told that they must take care of them. The control subjects were given plants and told that the staff would water them. The responsibility-induced group became more active, showed better mood, and had fewer health declines. In a subsequent analysis, intervention subjects had lower mortality at 18 months (135).

COLLECTIVE EFFICACY INTERVENTIONS

This sixth category of interventions is distinguished because it targets larger community context rather than individuals or families. We refer to interventions undertaken at the neighborhood or

community level that seek to mobilize existing resources as collective efficacy interventions. We use this term broadly to refer to a diverse group of studies. Far fewer studies targeting community-level factors have been undertaken despite the obvious fit with social epidemiology theories. As Wallerstein and colleagues observe (136), while social epidemiologists and community-based interventionists have much in common, they have achieved only limited success at working together.

Several important themes can be seen in these studies. First, there has been an emphasis on community empowerment (137–140). Interest in the potential role of empowerment as a key psychosocial mediator linking adverse material conditions to poor health can be found throughout the landscape of social epidemiology (141, 142). Other models have sought to boost community resilience though the effective mobilization of locally available assets and leadership (143, 144). Models such as the Community Action Model (145) have been developed and refined to address health disparities in health-related behaviors such as tobacco, and can be generalized to a variety of other public health challenges.

Structural or multilevel interventions are of particularly high value and importance for the future of social epidemiology (1, 3, 34, 146, 147). While studies that can be considered structural are small in number, this approach is growing in popularity as social epidemiology has leaned toward multilevel strategies. The area that has received the most consistent attention has been HIV. Interventions that target both individual behaviors and community factors have been developed and tested to reduce disease transmission among sex workers (148) and in bathhouses (149) and for reducing homelessness (150) and gender-based violence (151). These complex interventions attempt to modify relevant risk regulators at the neighborhood level or within organizational settings (152).

Some have argued that complex multilevel interventions aimed at higher levels of organization cannot be evaluated using traditional randomized designs. In their review of interventions aimed at reducing health disparities in Holland, Stronks and Mackenbach (153) conclude that experimental or quasi-experimental designs can and have been effectively used in this area.

ORGANIZATIONAL AND JOB/SCHEDULE CONTROL INTERVENTIONS

Worksite interventions have been part and parcel of public health and occupational epidemiology since the mid-nineteenth century, as occupational health and safety programs were oriented to improving working conditions of mainly blue collar workers employed in industrialized settings exposed to toxic substances and other unsafe working conditions. Health protection efforts over the last century have reduced accidents, injuries, morbidity, and mortality. Over time, those interested in psychosocial intervention came to view worksites as excellent venues in which to conduct health promotion activities related to tobacco (154), alcohol consumption, diet (155), mental health (156), and physical activity (157). Organizational psychologists also became interested in worksite conditions with the aim of improving productivity in the workplace, reducing turnover and absenteeism, and improving mental health in employees. Over the last decade these groups have converged, leading to the first set of worksite interventions aimed at changing the organization of work to improve worker physical and mental health.

A landmark study by Theorell and colleagues provided crucial preliminary evidence (158). In a randomized experiment in a Swedish insurance company, they found that workers in a department where managers were trained to improve the psychosocial work environment had decreased cortisol levels compared to a control group. A meta-analysis of 48 experimental studies of the benefits of organization-focused interventions for work-related stress by van der Klink and colleagues (159) found small but significant effects on a range of outcomes from quality of life to psychological complaints. Effects were strongest for cognitive-behavioral interventions. More than a decade later, there are still few worksite intervention programs aimed at improving health by altering psychosocial features of organizational environments. There are signs that this is beginning to shift. A more recent review by Lamontagne suggests that job-stress interventions are becoming increasingly complex and focused on organization-level change (160). Organizational-level interventions have been launched in Holland (161) and elsewhere. Implementing complex interventions in workplaces can be especially challenging, and requires particular attention to the careful reporting of intervention process and fidelity (162). Next, we discuss two major initiatives that have been launched in the last several years that may well change the landscape of field experiments in social epidemiology.

The first of these is the Work, Family and Health Network, jointly sponsored by the NIH, Centers for Disease Control (CDC), and several foundations.[2] The second major effort is the NIOSH-sponsored WorkLife Initiative, which is part of the Total Worker Health Program (TWH). Both initiatives have launched major group randomized trials that are currently reaching their conclusions. Here we do not report on the results of these trials because they are not yet known, but we describe the development and initial aims of such interventions. The WorkLife Initiative (WLI) reflects a strategy for integrating occupational health and safety and health protection with health promotion to protect worker injury and illness and to advance health and well-being (163, 164). Issues related to the health promotion aspect of this program are discussed in great detail in Chapter 5. The Work, Family and Health Network (WFHN) is implementing a multisite randomized field experiment of workplace practices and policies on work-family life on health and other outcomes aimed at improving not only worker health but the health of family members and the productivity of the company. These two initiatives are innovative in being theory-driven, transdisciplinary, focused on organizational structure, and adhering to a rigorous experimental design. Here we describe each of them with some of the background studies that led up to the launching of larger group randomized trials.

THE WORK, FAMILY AND HEALTH NETWORK

The Work, Family and Health Network was started with an initial pilot phase, in which participating centers conducted studies to assess the viability of developing an intervention and assessing biomarkers and other indicators of health and well-being in employees and their families, as well as organization-level outcomes such as job turn-over and productivity. The findings suggested that managers' practices and attitudes concerning work-family life were associated with a host of cardiovascular risk factors including blood pressure, glycosylated hemoglobin, tobacco use, body

2 https://www.nichd.nih.gov/research/supported/Pages/workhealthinit.aspx

mass index (BMI), and cholesterol as well as sleep. Employees who worked for less supportive supervisors slept on average 29 minutes less than those with more supportive supervisors and had about twice the cardiovascular risk as their counterparts (165). Other centers developed promising interventions along two lines: One center developed a worksite cluster intervention to increase schedule control among employees by empowering them to redesign jobs. The intervention, called Results Only Work Environment (ROWE), showed improvements in health-related outcomes including improved sleep, and more appropriate help-seeking behavior (166). These intervention effects were in part mediated by changes in schedule control and negative work-home spillover. A third center developed and refined an intervention and related measures of family-supportive supervisor behaviors (167). In their initial study in grocery stores, among employees with high work-to-family conflict, interventions aimed at supervisors positively influenced job satisfaction, turnover intentions, and self-reported health. A fourth center examined work-family strain on children and family members, and found that when workers had low flexibility there was evidence of stressor transmission to their children (168).

The results from the initial phase of the network led to the design of a large-scale randomized group intervention in two large companies, one an IT company and the other a long-term care provider. The WFHN integrated the strongest elements from phase one interventions with components designed to increase schedule control, improve work-family life, and train managers in family-friendly supervisory behaviors. The intervention included participatory work redesign activities that identified new work practices and processes to increase employees' control over work time, while still meeting business needs, as well as supervisory training about strategies to demonstrate support for employees' personal and family lives, while also supporting job performance. In both industries, the intervention was conducted over six months and used a cluster group design with departments or facilities being randomized to intervention or to control. There are a number of family, corporate, and employee health outcomes in the study (169), but employee health outcomes centered on cardiometabolic health (blood pressure, HbA1c, BMI, cholesterol, tobacco consumption) combined into a modified Framingham risk factor score. Sleep was also assessed from actigraphy. At this time, results are not yet published, but these studies represent a new wave of organizational and psychosocial workplace interventions to improve employee, family, and corporate health.

THE WORKLIFE INITIATIVE

NIOSH established the WorkLife Initiative in 2004 to promote information, dissemination, research, and policy development relevant to the integration of worksite health protection and health promotion programs and policies.[3] Like the WFHN, a large component of this initiative was a demonstration project to guide further development of integrated employment-based healthcare programs. Some of these interventions are still in the field and results not yet in. However, early indications from related interventions suggest that the social context of worksites shapes employee behavior change (170). Many of these approaches are discussed more fully in Chapter 10. We mention them here because the lines between behavioral interventions and

3 http://www.cdc.gov/niosh/docket/archive/docket132.html

psychosocial interventions become increasingly blurred, especially in worksites where interventions often include organization-level efforts to reduce job strain.

Having provided a brief and selective review of seven types of intervention studies, we now turn to a review of the five main propositions suggested in the introduction. A more comprehensive list of studies and their characteristics and findings can be found in Table 11.1.

PROPOSITION 1: EXPLICATE THE THEORY

Good intervention designs require strong theory. Nevertheless, many psychosocial interventions are designed and evaluated in the absence of a clearly articulated theoretical platform. Theory can be important on three distinct levels as it relates to intervention design. First, grand-scale metatheories provide a big-picture framework including assumptions, sentinel concepts, and epistemological principles that guide the intervention at the highest level of abstraction. Second, middle-level theoretical models are useful in guiding the design of the intervention by providing a narrative account of how the intervention is supposed to work. These midlevel theoretical models also provide guidance about the optimal timing, intensity, and duration of intervention. Three examples of midlevel theoretical approaches will be discussed: social learning theory (SLT), transtheoretical models (TTM), and the social contextual model (SCM) of behavior change. A third level of theory provides disease-specific insights into natural history, framing the role that specific psychosocial mechanisms play at each point in the disease course. A smaller scale, disease-specific theory can anchor interventions to specific knowledge of pathophysiology and lifecourse dynamics that provide insights about the psychosocial trajectory associated with each disease. For example, interventions designed for diseases of sudden onset require different strategies compared with those for diseases of diffuse onset (e.g., heart attack vs. arthritis). Illnesses that resolve in recovery are likewise different from those that are chronically degenerative (e.g., stroke vs. multiple sclerosis). A sound intervention design can be improved by attending to all three levels of theory. Each is discussed briefly in turn.

METATHEORETICAL APPROACHES

In a classic article, the social anthropologist Roy D'Andrade proposed three tiers of general theory for three distinct domains of scientific inquiry (171). The first is the domain of the *physical sciences*, in which a limited set of basic objects and forces operate deterministically, and can be described in mathematical form using a limited set of laws that apply at all times and places. Second is the domain of *natural sciences*, including complex ecological, meteorological, and biological systems. In contrast to physical sciences, this tier involves the explication of the components, levels, and dynamics that arise in complex systems not governed by universal laws. Instead, general systems propositions are stated in probabilistic rather than deterministic form, as "natural language statements" that describe the tendencies that generally govern the behavior of complex systems. Finally, D'Andrade refers to the third domain

as *semiotic sciences*, in which relations arise not from finite universal laws or from general system tendencies, but are the consequence of sentient actors engaged in a world of meaning they themselves help to create. With its root in the Greek *sema*, meaning sign or symbol, the semiotic sciences embrace the importance, complexity, and unpredictability of consciousness, agency, and action.

The field of social epidemiology straddles two worlds. On one hand, we adopt a biomedical orientation to disease etiology grounded in the physical and natural sciences, in which experimental designs help to limit the interference that arises from human consciousness in evaluation of treatments. On the other hand, we embrace a more humanistic conception rooted in semiotic sciences like psychology and sociology, in which we seek to harness the power of symbols to improve health. While it may be possible to test and evaluate pharmacologic therapies using experimental designs premised on finding the universal causal effect of a biochemical process, psychosocial interventions require a semiotic science that embraces the complexities that arise from conscious actors. The mission of social epidemiology requires an understanding of humans both as biological entities and as actors that make meaning in a semiotic system that does not obey universal laws. Attempts have been made to formulate general systems theories that bridge the natural, physical, and semiotic sciences across multiple levels (172–177). To achieve legitimacy and acceptance within the dominant biomedical world, the natural science perspective remains tempting. At the same time, social epidemiologists must recognize the limits of treating psychosocial phenomena as though governed by finite and universal laws. This tension is acute in the study of psychosocial intervention. We borrow features from the natural and physical sciences, but must also account for the complexities of sentient human actors.

SYSTEMS THEORY: A TOOL FOR INTERVENTION

Systems science has begun to make significant inroads in epidemiology (3, 146, 178–181); however, these approaches remain controversial. While experimental trials remain the most compelling evidence for the efficacy of interventions, systems theories have played an important role in guiding health promotion strategies, particularly in the area of tobacco control (182–185). Increasingly, agent-based and system-dynamic models are being used in the planning and design of interventions to facilitate diffusion of behavior change across social network ties (186). Large-scale policy interventions are often the most powerful strategies to effect health improvement in whole populations. However, well-intentioned policy changes often yield unanticipated consequences that undermine gains in population health (187). Systems theories can help investigators anticipate and quantify the unintended consequences of interventions across levels. Systems theories and models may be particularly useful in understanding, predicting, and harnessing the social network dynamics that amplify or undermine interventions (188, 189). The use of systems theories can be helpful in the planning stages of an intervention study, during which simulations can be used to develop and refine theories about how an intervention at one level will have cascading influences at other levels. In the area of obesity, for example, systems theories are being used to improve the efficiency of intervention delivery (189–191).

MIDLEVEL THEORETICAL APPROACHES

SOCIAL LEARNING THEORY

The most common theory found in the field is the general family of cognitive-behavioral theories, best exemplified by social learning theory (SLT), described in Bandura's seminal work on the concept of self-efficacy (192). In this model, self-efficacy beliefs are shaped by behavior, as well as by feedback received from the social environment. Through observation of others, verbal support, and persuasion, self-efficacy beliefs form as domain-specific cognitive structures linked to health-promoting behaviors and to general well-being. Social learning theory principles have been employed in all intervention types described above. For example, SLT has informed the development of self-management programs in chronic illness (31, 193–196), and was the central theoretical model guiding the Stanford Five-City Project (a behavior change intervention) (25) and the more recent ENRICHD trial, as well as numerous social support interventions (197, 198). A fine discussion of theoretical issues in social learning theory and intervention design is found in Clark (198).

TRANSTHEORETICAL MODELS

The transtheoretical model (TTM) has emerged as a dominant theoretical orientation in the study of behavior change (199–201). In this model, behavioral change unfolds in a sequence of stages from precontemplation, to contemplation, to preparation, to action, and finally to maintenance (201). The model also proposes that different self-change strategies, or processes of change, are involved in moving between stages, and that different beliefs are associated with each stage. Through application to behavior change ranging from smoking cessation, to mammography screening, to condom use, TTM proponents have identified general patterns that can be used to optimize the timing and content of interventions. The model is *trans*theoretical because it assumes that no single theory can account for the complexities of behavior change. A key feature of what Prochaska calls the "stage paradigm" is that different intervention strategies are needed at each stage of behavior change. This concept of targeting the intervention to the stage of readiness for change has been a guiding principle in over 1,000 studies ranging from smoking cessation (202–205), to diet and physical activity (206–212), disease management (213–215), and reducing violence and injury (216). Recently, several commentators including Prochaska have raised concerns about limitations of TTM (217–221). Among the staunchest critics, West has argued that the model is so flawed that it is holding back progress in health promotion and should be scrapped (222). Some studies have found that, when measured, core variables from the model add little value for predicting behavior change (223). Many argue that there is little evidence that stage-based interventions are more effective (224). TTM may not be an adequate tool for understanding, for example, the role of social network and contagion in behavior change. Future work may benefit from a hybridization of TTM with other models that can better incorporate diverse, nonpsychological environmental features that may be modifiable.

SOCIAL CONTEXTUAL MODEL OF BEHAVIOR CHANGE

Amid the frequent calls for interventions that focus on the social context in which behavior forms, a relatively new theoretical model, the social contextual model (SCM) of behavior change, has emerged. The model was developed by researchers at the Dana-Farber Cancer Institute and the Harvard School of Public Health to guide cancer prevention interventions (225, 226). The model, grounded in a rich foundation of behavioral science, delineates a sequence of psychosocial factors including self-efficacy, attitudes and beliefs, and intentions that underlie key health behaviors that are the targets of change. Where the model goes beyond other approaches is in making explicit the larger social structural characteristics through which psychosocial processes act as mediators. Attention is focused on the "modifying conditions" within the social context that are amenable to change and that can substantially alter the impact of behavior change efforts. This model has been successfully implemented in the United States to reduce tobacco use (227) and to increase fruit and vegetable intake (170). Recently, Nagler and colleagues have demonstrated a step-by-step process for implementing this model for smoking cessation within schools in India (228). The SCM approach represents an important advance in building theory-based interventions that explicitly deal with larger structural contexts.

DISEASE-SPECIFIC THEORIES

The final level of theory is perhaps the smallest in scale, and pertains to disease-specific considerations of how the natural history of a disease relates to particular psychosocial risk factors or behavioral inputs. A disease-specific theory is an essential component of intervention design because just as any disease has a natural history at a physiologic level, so too do diseases follow a psychosocial and behavioral trajectory. In the absence of an understanding of the sequence of psychosocial issues as they relate to disease etiology or course, interventions risk being improperly timed and sequenced (see 229 for a review). One example of this type of theory has been type-A behavior (TAB) and cardiovascular disease (41, 230–233). After initial enthusiasm, interest in TAB has waned since the 1980s (234). However, TAB remains a good example of the usefulness of small-scale theory because it is grounded in a more general theoretical model (cognitive behavioral theory), draws on knowledge of the natural history of a disease (atherosclerosis), and yields specific hypotheses that link psychosocial mechanisms (hostility, impatience, and time-urgency) to particular physiologic pathways (overstimulation of sympathetic nervous system leading to elevated catecholamines and corticosteroids).

PROPOSITION 2: TARGET A SPECIFIC PSYCHOSOCIAL MECHANISM

The essential element of a psychosocial intervention is the systematic attempt to modify a psychosocial factor known to be associated with the desired outcome. This implies that intervention

design and testing should follow from earlier observational studies that document the relevance of some specified set of psychosocial factors that mediate or moderate the etiological process. As noted earlier, the development of a coherent theoretical model that provides a rationale for these links is a prerequisite. Greater specificity and detail regarding the proposed mechanism of action will lead to results with more meaning and interpretability. While observational studies can be hypothesis-generating or exploratory, *an intervention study must be hypothesis-driven*. The landscape of psychosocial intervention is filled with studies that succeeded in modifying some psychosocial process but failed to alter the intended disease outcome. In some cases, the intervention may have been too weak to alter psychosocial processes, or those processes are not sufficiently linked causally to the health outcome in question.

One common weakness is the failure to clearly specify the psychosocial target. Some investigators have used a "shot-gun" approach, in which a wide range of conceptually unrelated interventions are designed to improve overall well-being, quality of life, or adjustment. An example is the randomized study by Cain et al. (235) in which women with gynecologic cancer were given "counseling designed in response to the needs of women with gynecologic cancer in reducing long-term psychosocial distress." Many community-based primary prevention trials designed to modify behaviors previously shown to be associated with risk of coronary artery disease were launched in the 1970s, including the WHO collaborative trials (27), the Oslo trial (33), the Goteborg trial (21), the Stanford Three- and Five-City trials (25, 26) and the MRFIT trial (236). As suggested previously, the results of these trials have been generally disappointing. It is clearer now that targeting high-risk individuals for behavior change interventions is not simple. As Rose has argued, attempts to modify health behaviors without addressing the larger context in which those behaviors arise runs the risk of blaming the victim and results in interventions that are, in his words, behaviorally inappropriate (237). Summarizing decades of research, Berkman notes: "Over and over again, we have learned that asking individuals to change behavior in the absence of a supportive social and economic context is very hard" (187, p. 547). Although these trials are considered the flagship intervention studies within the field of social epidemiology, the uneven results may have arisen from too much *epidemiology* and not enough *social*.

While many of these trials featured strong designs with ample power, most were not based on an explicit theoretical foundation. Often, lifestyle and behaviors have not been conceptualized within larger structural contexts. Rather, these trials illustrate the biomedicalization of behavior in which smoking, diet, drug use, and other risk behaviors are addressed within a narrowly individualistic frame (for a thoughtful elaboration of this point, see 238). Behaviors and lifestyles are products of particular social contexts (3). The brutal simplicity and power of experimental designs provides no guarantee of success in the absence of the right intervention given at the right time and in an effective manner.

PROPOSITION 3: IDENTIFY AN APPROPRIATE HEALTH OR FUNCTIONAL OUTCOME

In designing a psychosocial intervention study, the choice of a primary outcome and how it should be measured is critical. There are two important considerations. First: Is the outcome

viewed as relevant, well accepted, and reliably measured by the larger audience to which this study is directed? Second: Is this outcome likely to be a sensitive marker of the beneficial impact of the intervention?

RELEVANCE OF THE OUTCOME

It is an accepted canon of clinical trial research that while the intervention itself can be novel, the primary outcome measure used to judge its effectiveness cannot (239). However, many trials have used vague or poorly measured outcomes not regarded by the wider audience as valid, reliable, and relevant to healthcare decision-makers (e.g., perceived distress, well-being, or psychosocial adjustment). Measurement problems are well known for these outcomes (240). Other researchers have created new outcome measures for a trial; in the absence of strong evidence about the reliability and validity of an outcome, the use of new, untested measures is to be avoided.

The question of what should be measured is controversial. There is considerable debate about whether the primary outcome should be the "hard" health outcome, or the "softer" behavioral or psychosocial mechanisms themselves. For example, Prochaska has argued that interventions that measure behavior change as their primary outcome (e.g., smoking or condom use) are more compelling because the risky behavior in question may be implicated in many health outcomes (201). Although this argument is compelling, the interventions that have had the greatest impact, and are more likely to be implemented and sustained, are those that have been evaluated using a "hard" outcome involving some health state or functional status. Prominent examples include studies by Ornish (reduction in sclerotic plaque), Spiegel (longer survival), and Frasure-Smith (fewer recurrent MIs). Countless psychosocial intervention trials have demonstrated changes in mechanisms but have failed to demonstrate expected health benefits. As an arm of social epidemiology, intervention work certainly has room for well-done studies of both types. In either case, those planning future trials should carefully choose a well-accepted and previously validated outcome measure that captures the intended influence. It is likely that studies that demonstrate health impacts will remain more compelling for scientists and policymakers.

PROPOSITION 4: CALIBRATE THE INTERVENTION TO THE LIFECOURSE

We argue that inadequate attention has been paid to the importance of the timing of interventions in the context of the lifecourse. The lifecourse approach provides insight into why many psychosocial intervention trials may not have yielded the effects anticipated on the basis of observational studies (5). This is a complex issue that is beginning to be discussed (5, 241–243). In essence, we argue that psychosocial intervention trials should carefully consider how best to calibrate the delivery of the intervention to the lifecourse trajectory of the target psychosocial process *and* the associated health outcome of interest. This calibration should also consider the timing of the evaluation of intervention impact. Thus, there are three distinct issues that arise from etiologic periods. First, when can we change the trajectory of the

exposure in ways that will lead to alterations of the target outcome? Second, when will that change impact the health outcome we are hoping to change? Third, how long will it take us to observe this potential effect on the outcome of interest? We focus here on these key concepts from the lifecourse perspective that are now understood to be important for epidemiology in general, and are increasingly relevant for intervention studies. We emphasize that issues related to etiologic period overlap with questions about latency, sensitive periods, cumulative disadvantage, and acute and current impacts.

First, it is important that interventions are matched to the relevant etiologic period. The etiologic period has multiple implications. Most observational studies examine a risk factor at one period in time. In such studies, it is impossible to identify the etiologic period of risk. For some exposures, especially those related to tobacco consumption and selected cardiovascular risk factors, longitudinal information on exposures over long time periods does exist, and epidemiologists can start to identify etiologic periods of risk with some accuracy. In social epidemiology, we are just starting this endeavor. Lifecourse models identify where exposures may have the most important impacts. Three distinct models have been proposed to understand lifecourse implications and etiologic period (244–247). The first model, dominant in developmental studies, is related to critical or sensitive periods in which early childhood or even prenatal exposures shape subsequent outcomes that may or may not be evident for years. In this model, early exposures that occur during a finite window of vulnerability shape subsequent outcomes independently of later experiences or changes in exposure. The exposure may not lead to obvious outcomes until later life, owing to substantial latency in the impact of exposure. In the second lifecourse model, exposures throughout life have a cumulative effect. In such cases, sensitive periods are of diminished importance; rather, it is cumulative exposures over many years that have the largest impact. In the final lifecourse model, early exposures may shape opportunities or barriers that interact synergistically with critical exposures in later life, which are themselves linked to disease outcomes. In this model, effect heterogeneity arises due to early-life factors that create increased vulnerability to later exposures. This third model is often called a social trajectory (5). The implications of these models for psychosocial intervention are substantial. The first model suggests that interventions should be narrowly targeted at early-life exposures, before the critical window closes. Later interventions may not be effective at altering the negative consequences of early-life programming. The second model suggests that intervening in the cumulative consequences of chronic exposure might be important, but that interventions that take place later may be weak. The third model suggests the possibility that interventions can and should consider differential patterns of earlier exposures to account for differential vulnerability. This model provides insights into why in some psychosocial interventions benefits are not seen (or in some cases harm participants) from intervening on groups that may have had substantially harsher early-life exposures.

Often, psychosocial interventions are based on observational studies showing a (general) association between a psychosocial exposure and some outcome. For example, ENRICHD and FIRST interventions were motivated by studies showing that social networks and support are beneficial after stroke and heart attack. Neither intervention showed an overall benefit on the primary outcomes. In both cases, subgroup analyses revealed that the intervention was less effective or even harmful in frail subjects. This finding suggests that interventions that occur too late—after the etiologic period may have closed and multisystem dysregulation becomes entrenched—may undermine our ability to detect a benefit suggested by observational studies. On the other hand,

in some studies, the healthiest subjects and those who are very early in the etiologic period may be most likely to volunteer for a clinical trial, which can lead to Type II error if the intervention is effective but the outcomes are assessed before the true risk of disease can be observed. However, worksite interventions that succeed in changing outcomes indicate that current exposures can have relatively immediate effects without long latency periods. Positive effects can even be observed among workers who presumably had been exposed for long periods and have some level of cumulative disadvantage. These interventions in later life suggest that plasticity and resilience occur in various ways across the lifecourse.

A central contribution of the lifecourse approach has been the concept of critical or sensitive periods, in which early childhood or even prenatal exposures shape subsequent outcomes. As etiologic research on fetal programming unfolds, it suggests that interventions to alter trajectories of cardiovascular risk may be more impactful very early in life, but may not yield measurable benefits for many decades. While discussions of sensitive periods and other lifecourse phenomena are becoming commonplace, there is little consensus about which specific exposures are related to which lifecourse dynamics for which health outcomes (5). Moreover, if the stage for adult chronic disease is set in very early life (or even prenatally), this would seem to imply that psychosocial intervention in adulthood may be too late. However, for many diseases of interest, the etiologic period is no doubt long, and what matters for population health may be repeated or cumulative exposure to risk over the lifecourse. Observational studies are often not extremely helpful in distinguishing between cumulative exposures and sensitive periods, which can be as conflated as age-period-cohort effects (248). This suggests that interventions outside of a sensitive period may still alter the probability of disease in late life. Psychosocial intervention researchers should ideally identify periods of the lifecourse where the intervention is likely to be most impactful, while balancing the need to minimize the gap between administration of treatment and recording of its effects.

A third, critical concept from the lifecourse perspective relates to the time it takes for an intervention's efficacy to be fully realized. Social epidemiology leans toward changing upstream factors related to health through a temporally and spatially complex and distal chain of events that may take decades to unfold. The RCT design is best adapted to observing causal effects of discrete interventions with a short, and well-understood period of latency. Policy interventions with long latency periods and spillover effects may be best tested using pseudo-randomized retrospective designs and "big data."

PROPOSITION 5: STRIVE FOR THE STRONGEST POSSIBLE EXPERIMENTAL (OR QUASI-EXPERIMENTAL) DESIGN

Psychosocial intervention has a credibility problem. As illustrated by the so-called great debate in the journal *Psychosomatic Medicine* in 2002 (249–253), critics claim that there is "no good clinical evidence to reject the null hypothesis" that psychosocial interventions have no direct effect on health (252). It is hard to debate the implications: here is a need for the strongest and most

rigorous designs possible to evaluate the efficacy and effectiveness of psychosocial interventions. Many studies have been methodologically (if not theoretically) substandard. Thousands of studies have been done using samples too small to detect differences, nonrandomized and weak designs, vague and unreproducible interventions, and outcome measures that are either not well accepted or are insensitive to the influences of the treatment. The majority of these studies have no impact on clinical practice or public policy. Perhaps worse, many biomedical researchers and policymakers do not know that exquisitely done studies do exist.

We conclude that social epidemiologists cannot afford the luxury of weak study designs, despite the many challenges of conducting experimental designs for psychosocial interventions. Although, as will be argued, the RCT is not the only design for evaluation, it remains the most well accepted standard of intervention evaluation, and offers the most powerful evidence of both efficacy and etiological significance. Oakes goes furthest in arguing that "group-randomized designs are the canonical research design for social epidemiology" (254). In what follows, we lay out a selection of the most important methodological issues faced by investigators seeking to evaluate psychosocial interventions in an attempt to avoid the most common methodological shortcomings that lead to inconclusive, unconvincing, or uninterruptable results.

STANDARDIZATION OF INTERVENTIONS

In a classically designed RCT to test a pharmacological treatment, both the active agent and the placebo are standardized for content and for dosage. While the standardization of psychosocial interventions is obviously more difficult, it is no less important. Any intervention that involves communication or behavior is likely to vary both across subjects and across interventionists. Some would argue that the challenge of standardization of psychosocial interventions make the use of experimental designs impossible. We do not share this view; investigators are obligated to take steps in the design of the intervention to maximize the degree of standardization. The key to standardization is to develop a thorough and detailed intervention protocol and manual prior to launching the study, and a rigorous quality assurance program during the study. The intervention protocol describes the procedures and policies of the intervention a priori. Interventionists should be trained and tested to ensure that the intervention is being implemented reliably across interventionists, and careful real-time monitoring of adherence to the protocol should be put in place.

Investigators must balance the need for standardization with the possibility of an overly rigid and structured protocol. One approach to standardization is the use of scripted presentations. These are useful for educational interventions involving homogeneous populations, but may be overly structured in other settings. An alternative model has been developed for the Families in Recovery from Stroke Trial (FIRST). In this approach, we recognize the need to shape the content of the intervention to the particular psychosocial needs of the each family. Balance is achieved across these competing goals through the use of an instrument designed to record the content of individual intervention sessions. This instrument comprises a matrix, with each session in a column, and all of 16 primary content domains in the rows. Interventionists record the extent to which each content domain is discussed in each of 15 intervention sessions. In the early phase of the intervention, the content of the sessions is dictated by the needs of each stroke survivor. In later sessions, the interventionist addresses any content areas that have not been previously

discussed. Using this tool, it becomes possible to track the content of the intervention as delivered, ensuring that some attention is paid to all content domains, while at the same time allowing for the flexibility to tailor the intervention to the needs of families.

In summary, standardization of the intervention is an important issue that requires planning and balance. The benefits of standardization go beyond the technical virtues of study design. Previous reports suggest that more tightly structured interventions improve attendance and satisfaction (255). One promising new methodology for standardizing interventions is the use of computer-based expert-systems programming to create individually tailored informational and feedback interventions. This technique has been used extensively by Prochaska and colleagues in work on smoking cessation (256). Not only can expert-systems-based interventions be highly standardized using data-derived algorithms, but it becomes possible to record the content of the intervention sessions precisely.

BLINDING

The cardinal virtues of the RCT design are randomization and blinding (sometimes called masking). The former assures that any differences seen between the two groups after intervention can be used to estimate the causal effect of the intervention. Blinding is a procedure that further strengthens our confidence in the internal validity of the results by removing (or minimizing) potential sources of bias. In the typical pharmacological trial, the gold-standard practice is a triple-blinded study in which the subject, the investigator (the physician), and the evaluator are all blinded as to treatment status. In a psychosocial intervention, blinding the participant is normally both ethically and practically impossible. Psychosocial intervention requires the active engagement of the intellectual and emotional faculties of the participant. Efforts to approximate the RCT ideal of the blinded subject often backfire. We believe that the crucial issue in psychosocial intervention is how to ensure that the outcome assessor is blinded. In practice, it is very difficult to prevent the outcome assessor from becoming unblinded in the course of interacting with the subject, who is aware of treatment assignment. In the FIRST study, the project director called each subject prior to their follow-up interview to remind them not to reveal their treatment status. The interview includes a reminder to subjects not to make reference to any other study staff members they may have encountered. Despite working with elderly brain-injured patients, our experience is that these efforts are effective in minimizing the occurrence of unblinding. One general procedure that has been used in several other trials is to ask staff members who assess outcomes to guess the treatment status of each subject. Our experience in the FIRST study is that assessors do no better than chance. However, these data allow the investigators to quantify the effectiveness of blinding. In ENRICHD, outcome assessors were completely blinded with regard to treatment status when the outcome was a cardiovascular event or mortality (records were reviewed by parties without any knowledge of treatment status). When outcomes were based on self-report or interviewers' assessments of depression or function, interviewers could often make a better than average guess which treatment group the respondent was in. Thus, even though extensive efforts were made to keep the assessors blind to treatment status, participants often unwittingly reveal their status. In the WFHN study, interviewers were blinded to treatment status but often had informal ways of guessing to which arm

participants were randomized. Thus, blinding, while central to the RCT design, is not straightforward in psychosocial intervention.

THE SELECTION OF THE CONTROL GROUPS

In addition to random assignment, the strength of the clinical trial derives from the ability to compare the effects in the treatment group to a control condition. To the extent that the treatment and control groups are similar at baseline, differences between these groups are attributed to the effect of the intervention. In pharmacological trials, the control group is given a sham treatment to which they are blinded. In psychosocial intervention, sham treatments are infeasible for both technical and ethical reasons. Alternatively, four approaches have been used: *(1)* usual care (UC) controls, *(2)* attentional controls, *(3)* information-only controls, and *(4)* waiting-list controls. In the former model, care is taken to ensure that control group subjects receive identical medical and social services. This approach can be especially problematic in disease management interventions, in which the target of the intervention is often the removal of barriers to increased medical utilization. While it is nearly impossible to ensure that the two groups differ only on exposure to the intervention, the best approach to achieving a clean design is to institute a process assessment strategy in order to monitor health care access, differences in screening, and greater levels of attention. Primary care providers who are aware that a given patient is in the intervention arm of a trial may be more likely to follow a patient aggressively. For this reason, it is advisable to minimize communication between the intervention staff and usual care partners.

In order to ensure that treatment differences are not the result of the expectancy effects engendered by enhanced attention, some trials have employed attentional controls. This practice is more common in psychiatric trials. Attentional controls in psychosocial intervention are highly problematic in our view because they add considerable expense to the intervention design, and they are likely to backfire. It is very difficult to train an attentional control provider to remain inert in the course of interacting with ill patients. In the aftermath of serious illness, any attempt to engage patients and family in discussion that is not in some way therapeutic violates social norms and can damage rapport between the study and its participants. Control subjects who see through the blatant attempt to provide inert attentional controls are likely to withdraw from the study—a methodological disaster.

An additional approach to the control condition has been to use information-only controls. In this method, control subjects are given written educational materials or are provided with informational sessions by study staff. The latter example is a special case of an attentional control. While easy to standardize, the provision of information without a context or the opportunity to ask questions can be an empty gesture. Providing enriched educational control conditions risks biasing the results of the trial toward the null. The use of an educational control condition is a common feature of multi-arm trials in which the main intervention is tested against a usual care control and an educational control to see whether the addition of the main intervention is of benefit. This multicontrol design should be avoided in circumstances in which the addition of a separate treatment arm would reduce the power of the study to detect reasonably small effect sizes in the other two arms.

One final method is the use of the waiting-list control, in which the intervention is offered to volunteers on a first-come, first-served basis until the desired sample size is achieved (see for example 257). Subsequent volunteers are used as the control subjects. This design is used as a precursor to a cross-over design, in which the control subjects are offered the intervention at the end of the evaluation period. The use of waiting-list controls is controversial. If intervention timing is critical, this approach is clearly inferior to other models. Also, care must be taken to ensure that those who volunteer first are not systematically different in motivation, level of access to information about the study, or in illness severity as compared with later volunteers.

SUBJECT RECRUITMENT AND ENROLLMENT

One strength of the RCT is the extent to which internal validity is maximized by study design. The weakness is the significant threats that exist to external validity if those who volunteer are different in the risk of the outcome or sensitivity to the intervention compared with the target population of interest. Any clinical trial is, by definition, a study of volunteers who may be higher functioning, younger, less sick, less marginalized, more amendable to change, or in other important ways different from the population in which the intervention might conceivably be implemented. The limitations of volunteer subject recruitment are potentially substantial and understudied, especially in psychosocial interventions (258–263). The danger of randomizing healthier, better-educated subjects is that a potentially efficacious intervention will be shown to be ineffective if the outcome distribution in the control group is a biased representation of the outcome distribution in the larger population in the absence of the intervention (Type II error).

Subject recruitment often takes place within an institutional setting (e.g., a hospital, a worksite, a community clinic) or via community outreach. In the case of institutional recruitment, it is extremely important to standardize recruitment and screening procedures to document the process in order to address the possibility of selection bias. Care must be taken to ensure that screeners are not sifting through potential candidates looking for those who might benefit most from the intervention. Newspaper, radio, or organizational newsletters can be used to disseminate information about the availability of an intervention trial. This approach has limitations if these outlets fail to reach the population most likely to benefit. As illustrated by the experience of the Family Support Project, media outreach tends to select people who are either acutely in crisis or who have waited too long after the onset of a crisis for the intervention to be maximally effective (264). Expedient methods of boosting recruitment can also lead to heterogeneity in the timing of intervention start-up. The negative consequences of volunteer bias are thus heightened.

SUBGROUP DIFFERENCES IN OUTCOME: HETEROGENEOUS TREATMENT EFFECTS

The most intriguing and controversial analysis from ENRICHD is the subgroup analysis of outcomes looking at gender and racial/ethnic differences. In prespecified analyses, statistically

significant interactions between gender and intervention revealed that men in the intervention had better outcomes than did men in the UC group. Women in the intervention group had worse outcomes than did women in the UC group. These findings paralleled those from the Montreal Study. In post hoc analysis with stratification by both gender and race/ethnicity, white men were shown to benefit from the intervention (HR = 0.80, $p = 0.10$), whereas other groups showed no benefit from participating in the intervention (265). Furthermore, considering cardiovascular disease outcomes, white men in the treatment group had lower risk (0.63, $p = 0.004$) of reinfarction or cardiovascular mortality; white women and black men and women experienced no benefit from being in the intervention group. White men were more likely to be married and better educated, had the fewest chronic conditions, better ejection fractions, and less severe MIs, and were more likely to receive thrombolytic therapy and cardiac catheterization and coronary revascularization. None of these factors accounted for the difference in outcomes by treatment group seen in ENRICHD. However, such differences suggest that unmeasured covariates may account for these sources of effect heterogeneity.

In FIRST, an RCT designed to improve functioning in stroke patients based on social network intervention (61, 266), heterogeneous effects were also observed. The primary outcome was functional independence at 6 months post stroke. The results showed no difference between intervention and UC in the primary endpoint at either 3 or 6 months, using an intention-to-treat approach. An examination of prespecified subgroups revealed that those who were not depressed and had little cognitive impairment, more minor strokes, and fewer preexisting chronic conditions tended to benefit from the intervention. Those who were frailer tended to do better in the UC group compared with the intervention group. Post hoc subgroup differences were very revealing in this study as well. Using a frailty summary score based on these factors, nonfrail participants in the intervention group did better than did those in the control in functional outcomes ($p = 0.001$) and they had lower mortality rates ($p = 0.03$) (62). Among those who were frail, those in the UC group had better functional outcomes and mortality risk. The results of these two trials, coupled with the results from M-HEART, consistently demonstrated overall null results when using an intent-to-treat analysis. The intervention may have had more positive results in some subgroups, whereas in other subgroups, those in the UC group did better. More attention should be paid to identifying and recruiting subjects who stand the best chance of benefiting.

SUBJECT RETENTION AND FOLLOW-UP

Sample attrition is a major problem in clinical trial designs. Internal validity can only be assured if a high proportion of subjects who are randomized complete the trial and are included in the analysis. A strong design feature in this regard is the use of the intention-to-treat rule (ITT), in which all randomized patients are included in the analysis, regardless of their postrandomization status. This is a conservative but powerful strategy not often employed in psychosocial intervention studies to date (for further information, see 267, 268, 269).

While loss to follow-up is of general concern, even greater problems arise when rates of attrition differ by treatment status. Differential losses to follow-up may occur because subjects randomized to the control group are more likely to withdraw, feeling that they had been denied an exciting intervention program. Alternatively, subjects randomized to intervention may be challenged by the

intervention or may express their resistance to behavior change by withdrawal from the study. Either scenario is problematic because it decreases power and can induce sizable bias. Investigators must carefully consider the possibility of factors that might impact the probability of a subject completing the trial and adjust by including inflation factors to the power calculations. As an example, Mohide et al. (270) conducted a randomized trial of a family caregiver support program for the home management of patients with dementia. Although underpowered at the outset with only 30 subjects in the treatment and control group, the study was devastated by a loss to follow-up of 30%, due mostly to subjects being transferred to long-term care. This study exemplifies how the intervention itself can predispose subjects to withdraw in favor of alternative interventions.

REACTIVITY AND CONTAMINATION EFFECTS

More than pharmacological trials, tests of psychosocial interventions are susceptible to the influence of reactivity and contamination effects that can bias the results of the study. The risks of contamination are particularly acute when the randomization unit is some location in sociogeographic space (such as floors of a buildings, worksites, or even neighborhoods). If individuals in the intervention group have contact with subjects in the control group, interference effects can weaken efficacy. Although more difficult to detect, this may also produce demoralization in control subjects, leading to early withdrawal or nonadherence. Building insulation between intervention and control groups is an important aspect of intervention design when groups are to be the units of randomization. The influence of contamination from secular trends should also be considered. As has been argued, many public health interventions have been crippled because policy changes or natural trends overwhelm the potency of the intervention. Most importantly, it is crucial that power calculations should include realistic estimates of improvement among unexposed (control) groups.

Another seldom-appreciated aspect of reactivity effects is the impact of social comparison processes within support groups (58). In a support group context, subjects who judge themselves to be doing better than other members of the support group are likely to feel positively about the intervention. However, individuals who are doing worse may be negatively impacted, resulting in paradoxical effects. The same problem exists within behavioral training interventions when group education and feedback are used. At the same time, when the intervention is designed to target individuals at increased risk for a particular disease outcome, reactivity effects can occur if members of the control group change their behavior simply because they are labeled high risk by the screening process that found them eligible to participate. In the most notable example, meaningful evaluation of treatment benefits in the MRFIT trial was compromised because many of the men voluntarily reduced their risk factor exposure as a result of being labeled "high risk" by the study's recruitment procedures (271).

ALTERNATIVES TO CLINICAL TRIALS

While it has been argued that clinical trials offer the most compelling evidence in favor of the efficacy of psychosocial intervention, the clinical trial is not the only useful evaluation tool. One

of the most important criticisms of the RCT methodology is that marginalized groups tend to be excluded disproportionately from participation. This is a significant threat to the external validity of these studies. Minority groups are less likely to participate in RCTs, making alternative designs important to verify the usefulness of psychosocial interventions in diverse populations. Macintyre argues that public health has been resistant to the movement toward the use of RCTs based on the belief that behavioral and psychosocial interventions are fundamentally different from biomedical interventions delivered in the clinical setting (272). A number of alternative designs have merit and should be encouraged. Natural experiments have received increasing attention in social epidemiology (273–278). Investigators must balance the relative costs and complexities of the clinical trial design against those benefits. One lower-cost alternative is the use of single-group designs in particular populations, in which each subject is used as their own control. This design is useful in circumstances in which the outcome of interest is assumed to be rapidly responsive to intervention effects. This way, baseline, preintervention assessments can be assumed to be reliable and valid. An example is the home-based intervention to train caregivers to manage behavior problems in cognitively impaired elderly persons conducted by Pinkston and colleagues at the University of Chicago (279). This trial was solidly grounded in behavioral theory, and used preintervention interviews with both elderly clients and their caregivers as control conditions. Although improvements in the targeted behaviors were seen in 76 percent of the dyads, the absence of a distinct control group limits the internal validity of this design.

SUMMARY

To summarize this discussion of methodological issues, one might profitably ask: what are the things most likely to go wrong in a psychosocial intervention trial? Having reviewed hundreds of studies in the preparation of this chapter, there are three general methodological flaws that have, more than all others, led to poor research and inconclusive findings. The leading cause of faulty methodology is the use of underpowered tests. Of the dozens of intervention trials that report negative or inconclusive findings, the vast majority failed to adequately consider the sample sizes that would be required to detect effect sizes of the magnitude one would expect to encounter. More than half the studies reviewed had sample sizes of fewer than 30 cases per treatment condition. It is impossible to conclude much from trials of this size since they are so underpowered. Ironically, many of these trials are testing complex multi-arm treatment conditions without understanding that the complexity of the design has compromised the study's ability to adequately test even the first order hypothesis.

The second major methodological weakness of these studies is the selection of inappropriate or inconsequential outcome measures. In this era of managed care, a golden opportunity exists for social epidemiologists to develop and test low-cost, minimally invasive interventions that have real consequences for health and well-being. To the extent that researchers fail to pick outcome measures that are relevant to larger policy and practice questions, or pick outcomes that are not sensitive to the impact of the intervention planned, negative results often result.

Finally, many high-visibility intervention trials have failed to take account of the substantial complexities associated with changing behavior. In part, this is due to a lack of a coherent conceptual grounding, which places health behaviors within a broader biopsychosocial context. In other

cases, it involves a failure to maintain the intervention duration or intensity sufficiently to allow the flower of behavior change to bloom (another point made by 29). And in still other cases, it has resulted from a failure to adequately account for the contaminating influences of secular trends and cross-over effects.

CONCLUSION AND FUTURE RESEARCH DIRECTION

The field of psychosocial intervention remains in an awkward stage of growth. On one hand, many individual-level psychosocial interventions have produced mixed or negative results. In their review of the best trials of primary prevention of coronary heart disease (including both psychosocial and pharmacological interventions) McCormick and Skrabanek (280) conclude that no improvement in total mortality has been achieved. In this review, we have attempted to address some of the most important possible explanations for this disappointing track record, and point in the direction of strategies for improvement. There are other possibilities we have not considered, including heterogeneous treatment effects (for which secondary analyses provided some support), and the chance that the true causal agent was not targeted (and the results of observational studies linking social integration with cardiovascular health were due to confounding or reverse causation) (241). However, experimental evaluation of the efficacy of psychosocial intervention remains a crucial component of the portfolio of social epidemiology. More than ever, documenting inequalities is not enough. To achieve what Galea calls a consequentialist epidemiology (281), we consider new approaches to develop, test, and disseminate interventions that are practical and effective, while maintaining our commitment to strong design and valid outcome assessment. Investigators are becoming more sophisticated in the design and testing of psychosocial interventions, and new tools (the Internet, systems models) are becoming available, which raises hopes for the future. Moreover, the time is right for a reevaluation of methods and priorities. This chapter concludes with three observations about the future of psychosocial intervention research. First, we argue that intervention researchers must think in new ways about how to contextualize health behaviors by moving from individually focused interventions to interventions that are aimed at the social environments in which behaviors are lodged (including families, work-groups, neighborhoods, and communities). Second, rapid progress in wearable computing, the Internet, and connectivity are creating new avenues for delivering, evaluating, and disseminating psychosocial interventions. Third, we argue that maintaining the highest possible standards of methodology remains the optimal strategy to ensure forward movement.

THINKING BEYOND THE INDIVIDUAL

In light of inconsistent results from individually targeted interventions, leading figures have called for a shift in emphasis toward interventions focused on higher-level, upstream factors. McKinley called for a "new" public health, emphasizing population-level interventions (237, 282–285). Berkman emphasizes the need for a shift toward policies that channel social determinants (187).

Galea and Link recently identified social interventions as one of six critical paths toward progress for social epidemiology (276). In his framework for public health action, Frieden notes that interventions that address socioeconomic determinants of health, while more difficult, offer the greatest potential public health impact (286). A general trend can be seen toward health promotion efforts targeted at organizations, worksites, communities, and entire populations (For an excellent recent review, see 7). Health service researchers have shown repeatedly that experiments to modify public policies, particularly those that regulate what is and is not reimbursed, can have dramatic impact on health outcomes (for a review see 287).

Despite agreement about the importance of thinking beyond the individual, consensus has not yet been reached on how population-level and community-based interventions are best evaluated. Policy interventions may have impacts that are so broad that no comparison group can be established. Natural experiments have increased in frequency and rigor and offer useful models (275, 276, 288). Many of the issues that arise in the study of natural experiments are not new and can be addressed using traditional methods. As noted by Craig (288), quantitative natural experimental studies should only be attempted when groups subject to varying levels of exposure can be compared—using samples large enough to detect the expected effects—and when accurate data can be obtained on exposures, outcomes, and potential confounders. For example, Cerda and colleagues recently evaluated a natural experiment conducted in Medellin, Columbia, that compared community violence rates before and after construction of a cable car system (274). This study used propensity score matching to improve causal inferences drawn from comparisons between intervention communities and control communities.

Exemplary interventions targeting higher levels are increasing in numbers. In one example, community outreach was used to recruit African American women into a study to test a multimodal intervention to reduce HIV risk and to promote condom use. The intervention illustrates the importance of building gender and cultural competence into interventions that build on community assets (289). Another innovative example was conducted by Kelly and colleagues (290), in which popular opinion leaders were recruited and trained to modify community norms around HIV risk behaviors. This study moved beyond the focus on individuals by targeting community norms within venues. Other studies target structural interventions in workplaces. Berkman and colleagues are studying the impact of greater workplace flexibility on the health of employees (165).

Finally, there are several areas in which promising intervention models have been developed, but little evaluation research has been conducted. Examples include neighborhood interventions for violence prevention or helping to maintain the elderly in their homes. As epidemiologic evidence mounts as to the importance of diet and nutrition, innovative interventions aimed at communities and worksites will be needed. Promising examples include the 5-a-Day initiative at a community level (291) and the Working Well study aimed at worksites (292). While several useful policy interventions designed to reduce access barriers to underserved populations have been mounted, one additional area that appears promising is the use of community empowerment and social network mobilization concepts to reduce access barriers. Previous examples, including the Tenderloin Project, involving community mobilization among poor urban elderly persons, have not been well evaluated with respect to health-state outcomes (293). As observational studies accumulate, and more is known about the ways in which social environments

and other upstream factors shape the psychosocial processes that influence health, intervention models will be further enriched.

PSYCHOSOCIAL INTERVENTION IN THE ELECTRONIC AGE

One exciting development in psychosocial intervention has been the introduction of computers, sensors, and Web-based components. Investigators are increasingly using the Internet as a tool to reach a target audience more efficiently, and to deliver interventions that are easier to measure and standardize. Examples to date include substance use disorders and smoking (294–297), obesity and physical activity promotion (298–300), chronic disease (101, 301–303), breast cancer (304), and dementia (305, 306). Web-based interventions allow more detailed tailoring of interventions to specific subject characteristics (297). The Internet also allows coping and disease management interventions to be enhanced by the social support that can arise from subjects interacting virtually in ways that create self-sustaining, user-directed networks. There is growing evidence that this aspect of Internet-based interventions is particularly powerful in conditions that require intensive self-management such as kidney disease (303). In a series of novel experiments, Centola has demonstrated that the Internet can be used to generate spontaneous diffusion of behavior change through social networks in ways that magnify interventions in communities (307, 308). Therefore, Web-based interventions offer not only the potential for improved interventions for individuals, but also the possibility of harnessing natural networks in structural interventions aimed at communities and organizations. One important concern is that many intervention studies can be criticized for targeting the highest-functioning members of communities. Language and cultural differences, homelessness, and poverty generate further barriers to implementation of psychosocial interventions in marginalized populations. Internet-based interventions may then exacerbate disparities in the reach and impact of public health measures.

STICKING TO WHAT WORKS

In conjunction with the call for a shift from individual- to community-level interventions has come a parallel challenge to expand the range of evaluation methodologies to include "interpretive" and "qualitative" designs (7, 284). McKinlay invokes the language of Thomas Kuhn, whose classic work *The Structure of Scientific Revolutions* (309), described the changes over time in scientific "paradigms." According to McKinlay, an essential part of the "new" paradigm of public health is a critique of positivism and a move to accept alternative systems of evidence (284). As Kuhn argued, in order for the cracks and fissures that exist within any paradigm to accumulate to critical mass (at which point they become the impetus for a paradigm revolution) they must be articulated in a language that is common and acceptable to those on both sides of the paradigmatic divide. Moreover, if a new "paradigm," characterized by an increasing awareness of extraindividual, upstream factors in health is in ascendance, the use of "interpretive" and "qualitative" designs may be precariously premature. For these reasons, it has been a central theme of this chapter that psychosocial interventions, while

attempting to identify the impact of social and psychological phenomena on disease etiology, must nevertheless recognize the practical necessity of methodological overachieving within the framework of a noncontroversial science of evaluation. Otherwise, we are doing little more than "preaching to the choir."

REFERENCES

1. Glass TA, Goodman SN, Hernán MA, Samet JM. Causal inference in public health. Annu Rev Public Health. 2013;34:61–75.
2. Hernan MA. A definition of causal effect for epidemiological research. J Epidemiol Community Health. 2004;58(4):265–71.
3. Glass TA, McAtee MJ. Behavioral science at the crossroads in public health: extending horizons, envisioning the future. Soc Sci Med. 2006;62(7):1650–71.
4. Berkman LF. Seeing the forest and the trees—from observation to experiments in social epidemiology. Epidemiol Rev. 2004;26:2–6.
5. Berkman LF. Social epidemiology: social determinants of health in the United States: are we losing ground? Annu Rev Public Health. 2009;30:27–41.
6. Rose G. Sick individuals and sick populations. Int J Epidemiol. 2001;30(3):427–32; discussion 33–4.
7. Sorensen G, Emmons K, Hunt MK, Johnston D. Implications of the results of community intervention trials. Annu Rev Public Health. 1998;19:379–416.
8. Abrams DB, Emmons KM, Linnan LA. Health behavior and health education: the past, present, and future. In: Glanz K, Lewis FM, Rimer BK, editors. Health behavior and health education: theory, research, and practice. 2nd ed. San Francisco: Jossey-Bass; 1997. pp. 453–78.
9. Mullen PD, Mains DA, Velez R. A meta-analysis of controlled trials of cardiac patient education. Patient Educ Couns. 1992;19:143–62.
10. Orth-Gomér K, Schneiderman N, editors. Behavioral medicine approaches to cardiovascular disease prevention. Mahwah, NJ: Erlbaum; 1996.
11. Janssen V, De Gucht V, Dusseldorp E, Maes S. Lifestyle modification programmes for patients with coronary heart disease: a systematic review and meta-analysis of randomized controlled trials. Eur J Prev Cardiol. 2013;20(4):620–40.
12. Pischke CR, Scherwitz L, Weidner G, Ornish D. Long-term effects of lifestyle changes on well-being and cardiac variables among coronary heart disease patients. Health Psychology. 2008;27:584–92.
13. Horton ES. Effects of lifestyle changes to reduce risks of diabetes and associated cardiovascular risks: results from large scale efficacy trials. Obesity. 2009;17(Suppl 3): S43–S8.
14. Angermayr L, Melchart D, Linde K. Multifactorial lifestyle interventions in the primary and secondary prevention of cardiovascular disease and type 2 diabetes mellitus: a systematic review of randomized controlled trials. Ann Behav Med. 2010;40(1):49–64.
15. Wenger NK, Froelicher ES, Smith LK, Ades PA, Berra K, Blumenthal JA, et al. Cardiac rehabilitation: clinical practice guideline No. 17. Rockville, MD: Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research and the National Heart Lung and Blood Institute, 1995 AHCPR Publication No. 96–0672.
16. Gyarfas I. Review of community intervention studies on cardiovascular risk factors. Clin Exp Hypertens A. 1992;14(1–2):223–37.

17. Burg M, Berkman L. Psychosocial intervention in coronary heart disease. In: Stansfeld SA, Marmot MG, editors. Stress and the heart: psychosocial pathways to coronary heart disease. Williston, VT: BMJ Books; 2002. pp. 278–93.

18. Razin AM. Psychosocial intervention in coronary artery disease: a review. Psychosom Med. 1982;44(4): 363–87.

19. Rozanski A, Blumenthal JA, Kaplan J. Impact of psychological factors on the pathogenesis of cardiovascular disease and implications for therapy. Circulation. 1999;99(16):2192–217.

20. Suls J. Anger and the heart: perspectives on cardiac risk, mechanisms and interventions. Prog Cardiovasc Dis. 2013;55(6):538–47.

21. Wilhelmsen L, Berglund G, Elmfeldt D, Tibblin G, Wedel H, Pennert K, et al. The multifactor primary prevention trial in Göteborg, Sweden. Eur Heart J. 1986;7(4):279–88.

22. Luepker RV, Murray DM, Jacobs DR, Jr., Mittelmark MB, Bracht N, Carlaw R, et al. Community education for cardiovascular disease prevention: risk factor changes in the Minnesota Heart Health Program. Am J Public Health. 1994;84(9):1383–93.

23. Luepker RV, Rastam L, Hannan PJ, Murray DM, Gray C, Baker WL, et al. Community education for cardiovascular disease prevention: morbidity and mortality results from the Minnesota Heart Health Program. Am J Epidemiol. 1996;144(4):351–62.

24. Carleton RA, Lasater TM, Assaf AR, Feldman HA, McKinlay S. The Pawtucket Heart Health Program: community changes in cardiovascular risk factors and projected disease risk. Am J Public Health. 1995;85(6):777–85.

25. Farquhar JW, Fortmann SP, Flora JA, Taylor CB, Haskell WL, Williams PT, et al. Effects of communitywide education on cardiovascular disease risk factors. the Stanford Five-City Project. JAMA. 1990;264(3):359–65.

26. Farquhar JW, Fortmann SP, Maccoby N, Haskell WL, Williams PT, Flora JA, et al. The Stanford Five-City Project: design and methods. Am J Epidemiol. 1985;122(2):323–34.

27. World Health Organization. Health for all 2000. Copenhagen: World Health Organization, Regional Office for Europe; 1986.

28. Anonymous. Multiple risk factor intervention trial: risk factor changes and mortality results. Multiple Risk Factor Intervention Trial Research Group. JAMA. 1982;248(12):1465–77.

29. Susser M. The tribulations of trials—intervention in communities. Am J Public Health. 1995;85(2): 156–8.

30. Coreil J, Levin JS, Jaco EG. Life style—an emergent concept in the sociomedical sciences. Cult Med Psychiatry. 1985;9(4):423–37.

31. MacLean DR. Theoretical rationale of community intervention for the prevention and control of cardiovascular disease. Health Rep. 1994;6(1):174–80.

32. Rockhill B. The privatization of risk. Am J Public Health. 2001;91(3):365–8.

33. Hjermann I. A randomized primary preventive trial in coronary heart disease: the Oslo study. Prev Med. 1983;12:181–4.

34. Huang TT, Drewnosksi A, Kumanyika S, Glass TA. A systems-oriented multilevel framework for addressing obesity in the 21st century. Prev Chronic Dis. 2009;6(3):A82.

35. Puska P, Salonen J, Nissinen A, Tuomilehto J. The North Karelia project. Prev Med. 1983;12(1):191–5.

36. McLaren L, Ghali LM, Lorenzetti D, Rock M. Out of context? Translating evidence from the North Karelia project over place and time. Health Educ Res. 2007;22(3):414–24.

37. Nissinen A, Tuomilehto J, Korhonen HJ, Piha T, Salonen JT, Puska P. Ten-year results of hypertension care in the community: follow-up of the North Karelia hypertension control program. Am J Epidemiol. 1988;127(3):488–99.

38. Salonen JT. Prevention of coronary heart disease in Finland—application of the population strategy. Ann Med. 1991;23(6):607–12.

39. Friedman M, Powell LH, Thoresen CE, Ulmer D, Price V, Gill JJ, et al. Effect of discontinuance of type A behavioral counseling on type A behavior and cardiac recurrence rate of post myocardial infarction patients. Am Heart J. 1987;114(3):483–90.

40. Powell LH, Shaker LA, Jones BA, Vaccarino LV, Thoresen CE, Pattillo JR. Psychosocial predictors of mortality in 83 women with premature acute myocardial infarction. Psychosom Med. 1993;55(5):426–33.

41. Friedman M, Thoresen CE, Gill JJ, Ulmer D, Thompson L, Powell L, et al. Feasibility of altering type A behavior pattern after myocardial infarction. Recurrent Coronary Prevention Project Study: methods, baseline results and preliminary findings. Circulation. 1982;66(1):83–92.

42. Friedman M, Thoresen CE, Gill JJ, Ulmer D, Powell LH, Price VA, et al. Alteration of type A behavior and its effect on cardiac recurrences in post myocardial infarction patients: summary results of the recurrent coronary prevention project. Am Heart J. 1986;112(4):653–65.

43. Gould KL, Ornish D, Scherwitz L, Brown S, Edens RP, Hess MJ, et al. Changes in myocardial perfusion abnormalities by positron emission tomography after long-term, intense risk factor modification. JAMA. 1995;274(11):894–901.

44. Ornish D, Scherwitz LW, Doody RS, Kesten D, McLanahan SM, Brown SE, et al. Effects of stress management training and dietary changes in treating ischemic heart disease. JAMA. 1983;249(1):54–9.

45. Ornish D, Brown SE, Scherwitz LW, Billings JH, Armstrong WT, Ports TA, et al. Can lifestyle changes reverse coronary heart disease? The Lifestyle Heart Trial. Lancet. 1990;336(8708):129–33.

46. Hogan BE, Linden W, Najarian B. Social support interventions: do they work? Clin Psychol Rev. 2002;22(3):383–442.

47. Heaney CA. Enhancing social support at the workplace: assessing the effects of the caregiver support program. Health Educ Q. 1991;18(4):477–94.

48. Gonzalez S, Steinglass P, Reiss D. Putting the illness in its place: discussion groups for families with chronic medical illnesses. Fam Process. 1989;28:69–87.

49. Glass T, Berkman LF, editors. The Families in Recovery from Stroke Trial (FIRST): study design and progress report. Presentation at the Annual Meetings of the Gerontological Society; 1998 November 19–23; Philadelphia, PA.

50. Galanter M. Network therapy for alcohol and drug abuse: a new approach in practice. New York: Basic Books; 1993.

51. Galanter M, Keller DS, Dermatis H. Network therapy for addiction: assessment of the clinical outcome of training. Am J Drug Alcohol Abuse. 1997;23(3):355–67.

52. Wasylenki D, James S, Clark C, Lewis J, Goering P, Gillies L. Clinical issues in social network therapy for clients with schizophrenia. Community Ment Health J. 1992;28(5):427–40.

53. Lehtinen K. Need-adapted treatment of schizophrenia: family interventions. Brit J Psychiat. 1994;23:89–96.

54. Garrison V. Support systems of schizophrenic and nonschizophrenic Puerto Rican migrant women in New York City. Schizophr Bull. 1978;4(4):561–96.

55. Gottlieb BH, editor. Marshalling social support. Beverly Hills, CA: Sage; 1988.

56. Biegel DE, Shore BK, Gordon E. Building support networks for the elderly: theory and applications. Lauffer A, Garvin C, editors. Beverly Hills, CA: Sage Publications; 1984.

57. Gottlieb B. Theory into practice: issues that surface in planning interventions which mobilize psupport. In: Sarason IG, Sarason BR, editors. Social support: theory, research, and application. The Hague: Martinus Nijhoff; 1985. pp. 417–37.

58. Lanza AF, Revenson TA. Social support interventions for rheumatoid arthritis patients: the cart before the horse? Health Educ Q. 1993;20(1):97–117.

59. Berkman LF, Blumenthal J, Burg M, Carney RM, Catellier D, Cowan MJ, et al. Effects of treating depression and low perceived social support on clinical events after myocardial infarction: the Enhancing Recovery in Coronary Heart Disease Patients (ENRICHD) Randomized Trial. JAMA. 2003;289(23):3106–16.

60. Friedland JF, McColl M. Social support intervention after stroke: Results of a randomized trial. Arch Phys Med Rehabil. 1992;73(June):573–81.

61. Glass TA, Berkman LF, Hiltunen EF, Furie K, Glymour MM, Fay ME, et al. The Families in Recovery from Stroke Trial (FIRST): primary study results. Psychosom Med. 2004;66(6):889–97.

62. Ertel KA, Glymour MM, Glass TA, Berkman LF. Frailty modifies effectiveness of psychosocial intervention in recovery from stroke. Clin Rehabil. 2007;21(6):511–22.

63. Mant J, Carter J, Wade DT, Winner S. Family support for stroke: a randomised controlled trial. Lancet. 2000;356(9232):808–13.

64. Eldred C, Sykes C. Psychosocial interventions for carers of survivors of stroke: a systematic review of interventions based on psychological principles and theoretical frameworks. Brit J Health Psych. 2008;13(3):563–81.

65. Lewycka S, Mwansambo C, Rosato M, Kazembe P, Phiri T, Mganga A, et al. Effect of women's groups and volunteer peer counselling on rates of mortality, morbidity, and health behaviours in mothers and children in rural Malawi (MaiMwana): a factorial, cluster-randomised controlled trial. Lancet. 2013;381(9879):1721–35.

66. Knudson KG, Spiegel TM, Furst DE. Outpatient educational program for rheumatoid arthritis patients. Patient Couns Health Educ. 1981;3:77–82.

67. Salinardi TC, Batra P, Roberts SB, Urban LE, Robinson LM, Pittas AG, et al. Lifestyle intervention reduces body weight and improves cardiometabolic risk factors in worksites. Am J Clin Nutr. 2013;97(4):667–76.

68. Rimmer JH, Rauworth A, Wang E, Heckerling PS, Gerber BS. A randomized controlled trial to increase physical activity and reduce obesity in a predominantly African American group of women with mobility disabilities and severe obesity. Prev Med. 2009;48(5):473–9.

69. Ackermann RT, Finch EA, Brizendine E, Zhou H, Marrero DG. Translating the Diabetes Prevention Program into the community: the DEPLOY pilot study. Am J Prev Med. 2008;35(4):357–63.

70. Lipscomb ER, Finch EA, Brizendine E, Saha CK, Hays LM, Ackermann RT. Reduced 10-year risk of coronary heart disease in patients who participated in a community-based diabetes prevention program: the DEPLOY pilot study. Diabetes Care. 2009;32(3):394–6.

71. Spiegel D, Bloom JR, Kraemer HC, Gottheil E. Effect of psychosocial treatment on survival of patients with metastatic breast cancer. Lancet. 1989;2:888–91.

72. Spiegel D, Morrow GR, Classen C, Raubertas R, Stott PB, Mudaliar N, et al. Group psychotherapy for recently diagnosed breast cancer patients: a multicenter feasibility study. Psychooncology. 1999;8(6):482–93.

73. Spiegel D, Butler LD, Giese-Davis J, Koopman C, Miller E, DiMiceli S, et al. Effects of supportive-expressive group therapy on survival of patients with metastatic breast cancer: a randomized prospective trial. Cancer. 2007;110(5):1130–8.

74. Andersen BL, Yang HC, Farrar WB, Golden-Kreutz DM, Emery CF, Thornton LM, et al. Psychologic intervention improves survival for breast cancer patients: a randomized clinical trial. Cancer. 2008;113(12):3450–8.

75. Belanoff JK, Sund B, Koopman C, Blasey C, Flamm J, Schatzberg AF, et al. A randomized trial of the efficacy of group therapy in changing viral load and CD4 counts in individuals living with HIV infection. Int J Psychiatry Med. 2005;35(4):349–62.

76. Spiegel D. Mind matters in cancer survival. Psychooncology. 2012;21(6):588–93.

77. Spiegel D. Mind matters in cancer survival. JAMA. 2011;305(5):502–3.

78. Goodwin PJ, Leszcz M, Ennis M, Koopmans J, Vincent L, Guther H, et al. The effect of group psychosocial support on survival in metastatic breast cancer. N Engl J Med. 2001;345(24):1719–26.

79. Preyde M, Synnott E. Psychosocial intervention for adults with cancer: a meta-analysis. J Evid Based Soc Work. 2009;6(4):321–47.
80. Bantum EO, Donovan K, Owen JE. A systematic review of outcomes associated with psychosocial interventions for women with breast cancer. J Clin Outcomes Manage. 2007;14(6):341–52.
81. Binger CM. Psychosocial intervention with the child cancer patient and family. Psychosomatics. 1984;25(12):899–902.
82. Chow E, Tsao MN, Harth T. Does psychosocial intervention improve survival in cancer? A meta-analysis. Palliat Med. 2004;18(1):25–31.
83. Dale HL, Adair PM, Humphris GM. Systematic review of post-treatment psychosocial and behaviour change interventions for men with cancer. Psychooncology. 2010;19(3):227–37.
84. Faul LA, Jacobsen PB. Psychosocial interventions for people with cancer. In: Baum A, Revenson TA, Singer J, editors. Handbook of health psychology. 2nd ed. New York: Psychology Press; 2012. pp. 697–715.
85. Gordon WA, et al. Efficacy of psychosocial intervention with cancer patients. J Consult Clin Psychol. 1980;48(6):743–59.
86. Kangas M, Bovbjerg DH, Montgomery GH. Cancer-related fatigue: a systematic and meta-analytic review of non-pharmacological therapies for cancer patients. Psychol Bull. 2008;134(5):700–41.
87. Lepore SJ, Coyne JC. Psychological interventions for distress in cancer patients: a review of reviews. Ann Behav Med. 2006;32(2):85–92.
88. Meyer TJ. Meta-analysis of controlled studies of psychosocial interventions with adult cancer patients. Thesis, Pennsylvania State University; 1992.
89. Meyler E, Guerin S, Kiernan G, Breatnach F. Review of family-based psychosocial interventions for childhood cancer. J Pediatr Psychol. 2010;35(10):1116–32.
90. Owen JE, Klapow JC, Hicken B, Tucker DC. Psychosocial interventions for cancer: review and analysis using a three-tiered outcomes model. Psychooncology. 2001;10(3):218–30.
91. Raingruber B. The effectiveness of psychosocial interventions with cancer patients: an integrative review of the literature (2006–2011). ISRN Nursing. 2011:1–27.
92. Watson M. Psychosocial intervention with cancer patients: a review. Psychol Med. 1983;13(4): 839–46.
93. Brown SA. Studies of educational interventions and outcomes in diabetic adults: a meta-analysis revisited. Patient Educ Couns. 1990;16(3):189–215.
94. Angermayr L, Melchart D, Linde K. Multifactorial lifestyle interventions in the primary and secondary prevention of cardiovascular disease and type 2 diabetes mellitus: a systematic review of randomized controlled trials. Ann Behav Med. 2010;40(1):49–64.
95. Heinrich E, Schaper NC, de Vries NK. Self-management interventions for type 2 diabetes: a systematic review. Eur Diabetes Nurs. 2010;7(2):71–6.
96. Horton ES. Effects of lifestyle changes to reduce risks of diabetes and associated cardiovascular risks: results from large scale efficacy trials. Obesity (Silver Spring). 2009;17(Suppl 3):S43–8.
97. Padgett D, Mumford E, Hynes M, Carter R. Meta-analysis of the effects of educational and psychosocial interventions on management of diabetes mellitus. J Clin Epidemiol. 1988;41(10):1007–30.
98. Plante WA, Lobato DJ. Psychosocial group interventions for children and adolescents with type 1 diabetes: the state of the literature. Child Health Care. 2008;37(2):93–111.
99. Rubin RR, Peyrot M. Psychosocial problems and interventions in diabetes: a review of the literature. Diabetes Care. 1992;15(11):1640–57.
100. Fekete EM, Antoni MH, Schneiderman N. Psychosocial and behavioral interventions for chronic medical conditions. Curr Opin Psychiatry. 2007;20(2):152–7.

101. Paul CL, Carey ML, Sanson-Fisher RW, Houlcroft LE, Turon HE. The impact of web-based approaches on psychosocial health in chronic physical and mental health conditions. Health Educ Res. 2013;28(3):450–71.

102. Sansom-Daly UM, Peate M, Wakefield CE, Bryant RA, Cohn RJ. A systematic review of psychological interventions for adolescents and young adults living with chronic illness. Health Psychol. 2012;31(3):380–93.

103. Richardson JL, Shelton DR, Krailo M, Levine AM. The effect of compliance with treatment on survival among patients with hematologic malignancies. J Clin Oncol. 1990;8(2):356–64.

104. Haffner S, Temprosa M, Crandall J, Fowler S, Goldberg R, Horton E, et al. Intensive lifestyle intervention or metformin on inflammation and coagulation in participants with impaired glucose tolerance. Diabetes. 2005;54(5):1566–72.

105. Ratner R, Goldberg R, Haffner S, Marcovina S, Orchard T, Fowler S, et al. Impact of intensive lifestyle and metformin therapy on cardiovascular disease risk factors in the diabetes prevention program. Diabetes Care. 2005;28(4):888–94.

106. Fawzy F, Cousins N, Fawzy N, Kemeny M, Elashoff R, Morton D. A structured psychiatric intervention for cancer patients: I. Changes over time in methods of coping and affective disturbance. Arch Gen Psychiatry. 1990;47(8):720–5.

107. Fawzy FI, Kemeny ME, Fawzy NW, Elashoff R, Morton D, Cousins N, et al. A structured psychiatric intervention for cancer patients: II. Changes over time in immunological measures. Arch Gen Psychiatry. 1990;47(8):729–35.

108. Fawzy FI, Fawzy NW, Hyun CS, Elashoff R, Guthrie D, Fahey JL, et al. Malignant melanoma: effects of an early structured psychiatric intervention, coping, and affective state on recurrence and survival 6 years later. Arch Gen Psychiatry. 1993;50(9):681–9.

109. Mulder CL, Antoni MH, Emmelkamp PM, Veugelers PJ, Sandfort TG, van de Vijver FA, et al. Psychosocial group intervention and the rate of decline of immunological parameters in asymptomatic HIV-infected homosexual men. Psychother Psychosom. 1995;63(3–4):185–92.

110. Lorig K, Lubeck D, Kraines RG, Seleznick M, Holman HR. Outcomes of self-help education for patients with arthritis. Arthritis Rheum. 1985;28(6):680–5.

111. Lorig K, Holman HR. Long-term outcomes of an arthritis self-management study: effects of reinforcement efforts. Soc Sci Med. 1989;29(2):221–4.

112. Oldenburg B, Perkins RJ, Andrews G. Controlled trial of psychological intervention in myocardial infarction. J Consult Clin Psychol. 1985;53(6):852–9.

113. Evans RL, Bishop DS, Matlock AL, Stranahan S, Smith GG, Halar EM. Family interaction and treatment adherence after stroke. Arch Phys Med Rehabil. 1987;68:513–7.

114. Tune LE, Lucas-Blaustein MJ, Rovner BW. Psychosocial interventions. In: Jarvik LF, Winograd CH, editors. Treatments for the Alzheimer patient: the long haul. New York: Springer; 1988. pp. 123–36.

115. Levy LL. Psychosocial intervention and dementia: I. State of the art, future directions. Occup Ther Ment Health. 1987;7(1):69–107.

116. Levy LL. Psychosocial intervention and dementia: II. The cognitive disability perspective. Occupational Therapy in Mental Health. 1987;7(4):13–36.

117. Toseland RW, Rossiter CM. Group interventions to support family caregivers: a review and analysis. Gerontologist. 1989;29(4):438–48.

118. Bourgeois MS, Schulz R, Burgio L. Interventions for caregivers of patients with Alzheimer's disease: a review and analysis of content, process, and outcomes. Int J Aging Human Dev. 1996;43(1):35–92.

119. Lawton MP, Brody EM, Saperstein AR. A controlled study of respite service for caregivers of Alzheimer's patients. Gerontologist. 1989;29(1):8–16.

120. Radojevic V, Nicassio PM, Weisman MH. Behavioral intervention with and without family support for rheumatoid arthritis. Behav Ther. 1992;23:13–30.

121. Dusseldorp E, van Elderen T, Maes S, Meulman J, Kraaij V. A meta-analysis of psychoeduational programs for coronary heart disease patients. Health Psychol. 1999;18(5):506–19.

122. Linden W, Stossel C, Maurice J. Psychosocial interventions for patients with coronary artery disease: a meta-analysis. Arch Intern Med. 1996;156(7):745–52.

123. Frasure-Smith N, Prince R. The ischemic heart disease life stress monitoring program: impact on mortality. Psychosom Med. 1985;47:431–45.

124. Frasure-Smith N, Prince R. Long-term follow-up of the Ischemic Heart Disease Life Stress Monitoring Program. Psychosom Med. 1989;51:485–513.

125. Frasure-Smith N, Lesperance F, Prince RH, Verrier P, Garber RA, Juneau M, et al. Randomised trial of home-based psychosocial nursing intervention for patients recovering from myocardial infarction. Lancet. 1997;350(9076):473–9.

126. Maunsell E, Brisson J, Deschenes L, Frasure-Smith N. Randomized trial of a psychologic distress screening program after breast cancer: effects on quality of life. J Clin Oncol. 1996;14(10):2747–55.

127. Sheps DS, Freedland KE, Golden RN, McMahon RP. ENRICHD and SADHART: implications for future biobehavioral intervention efforts. Psychosom Med. 2003;65(1):1–2.

128. Enhancing recovery in coronary heart disease patients (ENRICHD): study design and methods. The ENRICHD investigators. Am Heart J. 2000;139(1 Pt 1):1–9.

129. Enhancing recovery in coronary heart disease (ENRICHD): baseline characteristics. Am J Cardiol. 2001;88(3):316–22.

130. Carney RM, Rich MW, teVelde A, Saini J, Clark K, Jaffe AS. Major depressive disorder in coronary artery disease. Am J Cardiol. 1987;60:1273–5.

131. Frasure-Smith N, Lesperance F, Talajic M. Depression following myocardial infarction: impact on 6-month survival. JAMA. 1993;270(15):1819–25.

132. Krantz DS, Schulz R. A model of life crisis, control, and health outcomes: cardiac rehabilitation of relocation of the elderly. In: Baum A, Singer JE, editors. Advances in environmental pyschology: applications of personal control. 2nd ed. Hillsdale, NJ: Erlbaum; 1980. pp. 23–57.

133. Rodin J, Rennert K, Solomon SK. Intrinsic motivation for control: fact or fiction. In: Baum A, Singer JE, editors. Advances in environmental pyschology: applications of personal control. 2nd ed. Hillsdale, NJ: Erlbaum; 1980. pp. 131–48.

134. Klein RF, Kliner VA, Zipes DP, Troyer WG, Wallace AG. Transfer from a coronary care unit. Arch Intern Med. 1968;122:104–8.

135. Rodin J, Langer EJ. Long-term effects of a control-relevant intervention with the institutionalized aged. J Pers Soc Psychol. 1977;35(12):897–902.

136. Wallerstein NB, Yen IH, Syme SL. Integration of social epidemiology and community-engaged interventions to improve health equity. Am J Public Health. 2011;101(5):822–30.

137. Laverack G, Wallerstein N. Measuring community empowerment: a fresh look at organizational domains. Health Promot Int. 2001;16(2):179–85.

138. Wallerstein N. Empowerment to reduce health disparities. Scand J Public Health. 2002;59(Suppl):72–7.

139. Wallerstein N. Powerlessness, empowerment, and health: implications for health promotion programs. Am J Health Promot. 1992;6(3):197–205.

140. Wilson N, Minkler M, Dasho S, Wallerstein N, Martin AC. Getting to social action: the Youth Empowerment Strategies (YES!) project. Health Promot Pract. 2008;9(4):395–403.

141. Marmot M. Health in an unequal world. Lancet. 2006;368(9552):2081–94.

142. Sampson RJ, Morenoff JD, Earls F. Beyond social capital: spatial dynamics of collective efficacy for children. Am Sociol Rev. 1999;64(5):633–60.

143. Morton MJ, Lurie N. Community resilience and public health practice. Am J Public Health. 2013;103(7):1158–60.

144. Poortinga W. Community resilience and health: the role of bonding, bridging, and linking aspects of social capital. Health Place. 2012;18(2):286–95.

145. Hennessey Lavery S, Smith ML, Esparza AA, Hrushow A, Moore M, Reed DF. The community action model: a community-driven model designed to address disparities in health. Am J Public Health. 2005;95(4):611–6.

146. Galea S, Riddle M, Kaplan GA. Causal thinking and complex system approaches in epidemiology. Int J Epidemiol. 2010;39(1):97–106.

147. Kaplan GA. What's wrong with social epidemiology, and how can we make it better? Epidemiol Rev. 2004;26:124–35.

148. Kang D, Tao X, Liao M, Li J, Zhang N, Zhu X, et al. An integrated individual, community, and structural intervention to reduce HIV/STI risks among female sex workers in China. BMC Public Health. 2013;13:717.

149. Ko NY, Lee HC, Hung CC, Chang JL, Lee NY, Chang CM, et al. Effects of structural intervention on increasing condom availability and reducing risky sexual behaviours in gay bathhouse attendees. AIDS Care. 2009;21(12):1499–507.

150. Kidder DP, Wolitski RJ, Royal S, Aidala A, Courtenay-Quirk C, Holtgrave DR, et al. Access to housing as a structural intervention for homeless and unstably housed people living with HIV: rationale, methods, and implementation of the housing and health study. AIDS Behav. 2007;11(6 Suppl):149–61.

151. Pronyk PM, Hargreaves JR, Kim JC, Morison LA, Phetla G, Watts C, et al. Effect of a structural intervention for the prevention of intimate-partner violence and HIV in rural South Africa: a cluster randomised trial. Lancet. 2006;368(9551):1973–83.

152. Saunders RP, Evans AE, Kenison K, Workman L, Dowda M, Chu YH. Conceptualizing, implementing, and monitoring a structural health promotion intervention in an organizational setting. Health Promot Pract. 2013;14(3):343–53.

153. Stronks K, Mackenbach JP. Evaluating the effect of policies and interventions to address inequalities in health: lessons from a Dutch programme. Eur J Public Health. 2006;16(4):346–53.

154. Moher M, Hey K, Lancaster T. Workplace interventions for smoking cessation. Cochrane Database Syst Rev. 2003(2):CD003440.

155. Geaney F, Kelly C, Greiner BA, Harrington JM, Perry IJ, Beirne P. The effectiveness of workplace dietary modification interventions: a systematic review. Prev Med. 2013;57(5):438–47.

156. Corbiere M, Shen J, Rouleau M, Dewa CS. A systematic review of preventive interventions regarding mental health issues in organizations. Work. 2009;33(1):81–116.

157. To QG, Chen TT, Magnussen CG, To KG. Workplace physical activity interventions: a systematic review. Am J Health Promot. 2013;27(6):e113–23.

158. Theorell T, Emdad R, Arnetz B, Weingarten AM. Employee effects of an educational program for managers at an insurance company. Psychosom Med. 2001;63(5):724–33.

159. van der Klink JJ, Blonk RW, Schene AH, van Dijk FJ. The benefits of interventions for work-related stress. Am J Public Health. 2001;91(2):270–6.

160. Lamontagne AD, Keegel T, Louie AM, Ostry A, Landsbergis PA. A systematic review of the job-stress intervention evaluation literature, 1990–2005. Int J Occup Environ Health. 2007;13(3):268–80.

161. Schelvis RM, Oude Hengel KM, Wiezer NM, Blatter BM, van Genabeek JA, Bohlmeijer ET, et al. Design of the Bottom-up Innovation project—a participatory, primary preventive, organizational level intervention

on work-related stress and well-being for workers in Dutch vocational education. BMC Public Health. 2013;13:760.

162. Egan M, Bambra C, Petticrew M, Whitehead M. Reviewing evidence on complex social interventions: appraising implementation in systematic reviews of the health effects of organisational-level workplace interventions. J Epidemiol Community Health. 2009;63(1):4–11.
163. Cherniack M, Henning R, Merchant JA, Punnett L, Sorensen GR, Wagner G. Statement on national worklife priorities. Am J Ind Med. 2011;54(1):10–20.
164. Sorensen G, McLellan D, Dennerlein JT, Pronk NP, Allen JD, Boden LI, et al. Integration of health protection and health promotion: rationale, indicators, and metrics. J Occup Environ Med. 2013;55(12 Suppl):S12–8.
165. Berkman LF, Buxton O, Ertel K, Okechukwu C. Managers' practices related to work-family balance predict employee cardiovascular risk and sleep duration in extended care settings. J Occup Health Psychol. 2010;15(3):316–29.
166. Moen P, Kelly EL, Hill R. Does enhancing work-time control and flexibility reduce turnover? A naturally occurring experiment. Soc Probl. 2011;58(1):69–98.
167. Hammer LB, Kossek EE, Anger WK, Bodner T, Zimmerman KL. Clarifying work-family intervention processes: the roles of work-family conflict and family-supportive supervisor behaviors. J Appl Psychol. 2011;96(1):134–50.
168. Almeida DM, Davis KD. Workplace flexibility and daily stress processes in hotel employees and their children. Annals Am Acad Pol & Soc Sci. 2011;638(1):123–40.
169. Bray JW, Kelly EL, Hammer LB, Almeida, DM, Dearing JW, King RB, et al. An integrative, multilevel, and transdisciplinary research approach to challenges of work, family, and health. Research Triangle Park, NC: RTI International, 2013. MR-0024–1302 Contract No.: RTI Press publication No. MR-0024–1302.
170. Sorensen G, Stoddard AM, Dubowitz T, Barbeau EM, Bigby J, Emmons KM, et al. The influence of social context on changes in fruit and vegetable consumption: results of the healthy directions studies. Am J Public Health. 2007;97(7):1216–27.
171. D'Andrade R. Three scientific world views and the covering law model. In: Fiske DW, Shweder RS, editors. Metatheory in social science. Chicago, IL: University of Chicago Press; 1986. pp. 19–41.
172. Antonovsky A. Complexity, conflict, chaos, coherence, coercion and civility. Soc Sci Med. 1993;37(8):969–74.
173. Bar-Yam Y. Improving the effectiveness of health care and public health: a multiscale complex systems analysis. Am J Public Health. 2006;96(3):459–66.
174. Bertalanffy Lv. General systems theory. New York: George Braziller; 1969.
175. Bronfenbrenner U. Developmental ecology through space and time: a future perspective. In: Moen P, Elder GHJ, editors. Examining lives in context: perspectives on the ecology of human development. Washington, DC: American Psychological Association; 1995. pp. 619–47.
176. Bronfenbrenner U. Ecological systems theory. In: Vasta R, editor. Six theories of child development: revised formulations and current issues. London, England: Jessica Kingsley; 1992. pp. 187–249.
177. Meadows DH. Thinking in systems: a primer. Wright D, editor. White River Junction, VT: Chelsea Green Publishing; 2008.
178. Auchincloss AH, Diez Roux AV. A new tool for epidemiology: the usefulness of dynamic-agent models in understanding place effects on health. Am J Epidemiol. 2008;168(1):1–8.
179. Diez Roux AV. Complex systems thinking and current impasses in health disparities research. Am J Public Health. 2011;101(9):1627–34.
180. Koopman JS. Infection transmission science and models. Jpn J Infect Dis. 2005;58(6):S3–8.
181. Krieger N. Theories for social epidemiology in the 21st century: an ecosocial perspective. Int J Epidemiol. 2001;30(4):668–77.

182. Alebiosu CO, Ayodele OE. The global burden of chronic kidney disease and the way forward. Ethn Dis. 2005;15(3):418–23.

183. Bambra CL, Hillier FC, Moore HJ, Summerbell CD. Tackling inequalities in obesity: a protocol for a systematic review of the effectiveness of public health interventions at reducing socioeconomic inequalities in obesity amongst children. Syst Rev. 2012;1(1):16.

184. National Cancer Institute. How to anticipate change in tobacco control systems. In: Best A, Clark PI, Leischow SJ, Trochim WMK, editors. Greater than the sum: systems thinking in tobacco control. NIH Pub. No. 06–6085. Bethesda, MD: U.S. Department of Health and Human Services, National Cancer Institute; 2007. pp. 111–48.

185. Luke DA, Harris JK, Shelton S, Allen P, Carothers BJ, Mueller NB. Systems analysis of collaboration in 5 national tobacco control networks. Am J Public Health. 2010;100(7):1290–7.

186. Levy DT, Bauer JE, Lee H-R. Simulation modeling and tobacco control: creating more robust public health policies. Am J Public Health. 2006;96(3):494–8.

187. Berkman LF. Unintended consequences of social and economic policies for population health: towards a more intentional approach. Eur J Public Health. 2011;21(5):547–8.

188. El-Sayed AM, Scarborough P, Seemann L, Galea S. Social network analysis and agent-based modeling in social epidemiology. Epidemiol Perspect Innov. 2012;9(1):1.

189. Hammond RA, Dube L. A systems science perspective and transdisciplinary models for food and nutrition security. Proc Natl Acad Sci U S A. 2012;109(31):12356–63.

190. Levy DT, Mabry PL, Wang YC, Gortmaker S, Huang TT, Marsh T, et al. Simulation models of obesity: a review of the literature and implications for research and policy. Obes Rev. 2010.

191. Maglio PP, Mabry PL. Agent-based models and systems science approaches to public health. Am J Prev Med. 2011;40(3):392–4.

192. Bandura A. Self-efficacy mechanisms in human agency. Am Psychol. 1982;37:122–47.

193. Lorig K, Holman H. Arthritis self-management studies: a twelve-year review. Health Educ Q. 1993;20(1):17–28.

194. Wilson W, Pratt C. The impact of diabetes education and peer support upon weight and glycemic control of elderly persons with noninsulin dependent diabetes mellitus (NIDDM). Am J Public Health. 1987;77(5):634–5.

195. Clark NM, Rakowski W, Wheeler JR, Ostrander LD, Oden S, Keteyian S. Development of self-management education for elderly heart patients. Gerontologist. 1988;28(4):491–4.

196. Lorig K, Laurin J, Gines GE. Arthritis self-management: a five-year history of a patient education program. Nurs Clin North Am. 1984;19(4):637–45.

197. Clark NM, Janz NK, Becker MH, Schork MA, Wheeler J, Liang J, et al. Impact of self-management education on the functional health status of older adults with heart disease. Gerontologist. 1992;32(4):438–43.

198. Clark NM, Janz NK, Dodge JA, Sharpe PA. Self-regulation of health behavior: the "take PRIDE" program. Health Educ Q. 1992;19(3):341–54.

199. Abrams DB, Boutwell WB, Grizzle J, Heimendinger J, Sorensen G, Varnes J. Cancer control at the workplace: the Working Well Trial. Prev Med. 1994;23(1):15–27.

200. Painter JE, Borba CP, Hynes M, Mays D, Glanz K. The use of theory in health behavior research from 2000 to 2005: a systematic review. Ann Behav Med. 2008;35(3):358–62.

201. Prochaska JO, Redding CA, Evers KE. The transtheoretical model and stages of change. In: Glanz K, Lewis FM, Rimer BK, editors. Health behavior and health education: theory, research, and practice. 2nd ed. San Francisco, CA: Jossey-Bass; 1997. pp. 60–84.

202. Abdullah AS, Mak YW, Loke AY, Lam TH. Smoking cessation intervention in parents of young children: a randomised controlled trial. Addiction. 2005;100(11):1731–40.

203. Cabezas C, Advani M, Puente D, Rodriguez-Blanco T, Martin C. Effectiveness of a stepped primary care smoking cessation intervention: cluster randomized clinical trial (ISTAPS study). Addiction. 2011;106(9):1696–706.

204. Cole TK. Smoking cessation in the hospitalized patient using the transtheoretical model of behavior change. Heart Lung. 2001;30(2):148–58.

205. Lawrence T, Aveyard P, Evans O, Cheng KK. A cluster randomised controlled trial of smoking cessation in pregnant women comparing interventions based on the transtheoretical (stages of change) model to standard care. Tob Control. 2003;12(2):168–77.

206. King AC, Sallis JF, Dunn AL, Simons-Morton DG, Albright CA, Cohen S, et al. Overview of the Activity Counseling Trial (ACT) intervention for promoting physical activity in primary health care settings. Activity Counseling Trial Research Group. Med Sci Sports Exerc. 1998;30(7):1086–96.

207. Kirk AF, Higgins LA, Hughes AR, Fisher BM, Mutrie N, Hillis S, et al. A randomized, controlled trial to study the effect of exercise consultation on the promotion of physical activity in people with Type 2 diabetes: a pilot study. Diabet Med. 2001;18(11):877–82.

208. Kirk AF, Mutrie N, Macintyre PD, Fisher MB. Promoting and maintaining physical activity in people with type 2 diabetes. Am J Prev Med. 2004;27(4):289–96.

209. Peterson KE, Sorensen G, Pearson M, Hebert JR, Gottlieb BR, McCormick MC. Design of an intervention addressing multiple levels of influence on dietary and activity patterns of low-income, postpartum women. Health Educ Res. 2002;17(5):531–40.

210. Plotnikoff RC, Lippke S, Johnson ST, Courneya KS. Physical activity and stages of change: a longitudinal test in types 1 and 2 diabetes samples. Ann Behav Med. 2010;40(2):138–49.

211. Taymoori P, Niknami S, Berry T, Lubans D, Ghofranipour F, Kazemnejad A. A school-based randomized controlled trial to improve physical activity among Iranian high school girls. Int J Behav Nutr Phys Act. 2008;5:18.

212. Yoo JS, Hwang AR, Lee HC, Kim CJ. Development and validation of a computerized exercise intervention program for patients with type 2 diabetes mellitus in Korea. Yonsei Med J. 2003;44(5):892–904.

213. Chang L, McAlister AL, Taylor WC, Chan W. Behavioral change for blood pressure control among urban and rural adults in Taiwan. Health Promot Int. 2003;18(3):219–28.

214. Johnson SS, Driskell MM, Johnson JL, Prochaska JM, Zwick W, Prochaska JO. Efficacy of a transtheoretical model-based expert system for antihypertensive adherence. Dis Manag. 2006;9(5):291–301.

215. Suppan J. Using the transtheoretical approach to facilitate change in the heart failure population. Congest Heart Fail. 2001;7(3):151–5.

216. Burke JG, Denison JA, Gielen AC, McDonnell KA, O'Campo P. Ending intimate partner violence: an application of the transtheoretical model. Am J Health Behav. 2004;28(2):122–33.

217. Armitage CJ. Is there utility in the transtheoretical model? Br J Health Psychol. 2009;14(Pt 2): 195–210.

218. Callaghan RC, Taylor L. Mismatch in the transtheoretical model? Am J Addict. 2006;15(5):403.

219. De Vet E, De Nooijer J, De Vries NK, Brug J. Do the transtheoretical processes of change predict transitions in stages of change for fruit intake? Health Educ Behav. 2008;35(5):603–18.

220. Nigg CR, Geller KS, Motl RW, Horwath CC, Wertin KK, Dishman RK. A research agenda to examine the efficacy and relevance of the transtheoretical model for physical activity behavior. Psychol Sport Exerc. 2011;12(1):7–12.

221. Prochaska JO. Moving beyond the transtheoretical model. Addiction. 2006;101(6):768–74; author reply 74–8.

222. West R. Time for a change: putting the transtheoretical (stages of change) model to rest. Addiction. 2005;100(8):1036–9.

223. Carlson LE, Taenzer P, Koopmans J, Casebeer A. Predictive value of aspects of the transtheoretical model on smoking cessation in a community-based, large-group cognitive behavioral program. Addict Behav. 2003;28(4):725–40.

224. Diclemente CC. A premature obituary for the transtheoretical model: a response to West (2005). Addiction. 2005;100(8):1046–8; author reply 8–50.

225. Sorensen G, Barbeau E, Stoddard AM, Hunt MK, Kaphingst K, Wallace L. Promoting behavior change among working-class, multiethnic workers: results of the healthy directions—small business study. Am J Public Health. 2005;95(8):1389–95.

226. Sorensen G, Emmons K, Hunt MK, Barbeau E, Goldman R, Peterson K, et al. Model for incorporating social context in health behavior interventions: applications for cancer prevention for working-class, multiethnic populations. Prev Med. 2003;37(3):188–97.

227. Sorensen G, Stoddard A, Quintiliani L, Ebbeling C, Nagler E, Yang M, et al. Tobacco use cessation and weight management among motor freight workers: results of the gear up for health study. Cancer Causes Control. 2010;21(12):2113–22.

228. Nagler EM, Pednekar MS, Viswanath K, Sinha DN, Aghi MB, Pischke CR, et al. Designing in the social context: using the social contextual model of health behavior change to develop a tobacco control intervention for teachers in India. Health Educ Res. 2013;28(1):113–29.

229. Bloom JR, Kessler L. Risk and timing of counseling and support interventions for younger women with breast cancer. Journal of the National Cancer Institute Monographs. 1994(16):199–206.

230. Friedman M, Rosenman RH. Association of a specific overt behavior pattern with increase in blood cholesterol, blood clotting time, incidence of arcus senilis, and clinical coronary arterery disease. JAMA. 1959;112:653–65.

231. Nunes EV, Frank KA, Kornfeld DS. Psychological treatment for type-A behavior pattern: a meta-analysis of the literature. Psychosom Med. 1987;48:159–73.

232. Powell LH, Thoresen CE. Effects of type A behavioral counseling and severity of prior acute myocardial infarction on survival. Am J Cardiol. 1988;62(17):1159–63.

233. Smith TW, Anderson NB. Models of personality and disease: an interactional approach to Type A behavior and cardiovascular risk. J Pers Soc Psychol. 1986;50:1166–73.

234. Thoresen CE, Powell LH. Type A behavior pattern: new perspectives on theory, assessment, and intervention. J Consult Clin Psychol. 1992;60(4):595–604.

235. Cain EN, Kohorn EI, Quinlan DM, Latimer K, Schwartz PE. Psychosocial benefits of a cancer support group. Cancer. 1986;57:183–9.

236. Benfari RC. The Multiple Risk Factor Intervention Trial (MRFIT): III. The model for intervention. Prev Med. 1981;10:426–42.

237. Rose G. Sick individuals and sick populations. Int J Epidemiol. 1985;14(1):32–8.

238. McMichael AJ. Coronary heart disease: interplay between changing concepts of aetiology, risk distribution, and social strategies for prevention. Community Health Stud. 1989;13(1):5–13.

239. Levin JS, Glass TA, Kushi LH, Schuck JR, Steele L, Jonas WB. Quantitative methods in research on complementary and alternative medicine. A methodological manifesto. NIH Office of Alternative Medicine. Med Care. 1997;35(11):1079–94.

240. Mittelman MS. Psychosocial intervention research: challenges, strategies and measurement issues. Aging Ment Health. 2008;12(1):1–4.

241. Berkman LF, Ertel KA, Glymour MM. Aging and social intervention: life course perspectives. In: Robert H. Binstock and Linda K. George, editors. Handbook of aging and the social sciences. 7th ed. San Diego, CA: Elsevier Academic Press; 2011. pp. 337–51.

242. Olds DL, Sadler L, Kitzman H. Programs for parents of infants and toddlers: recent evidence from randomized trials. J Child Psychol Psychiatry. 2007;48(3–4):355–91.

243. Seguin M, Lesage A, Turecki G, Bouchard M, Chawky N, Tremblay N, et al. Life trajectories and burden of adversity: mapping the developmental profiles of suicide mortality. Psychol Med. 2007;37(11):1575–83.

244. Ben-Shlomo Y, Kuh D. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. Int J Epidemiol. 2002;31(2):285–93.

245. Hertzman C, Boyce T. How experience gets under the skin to create gradients in developmental health. Annu Rev Public Health. 2010;31:329–47, 3p following 47.

246. Hertzman C, Power C. Health and human development: understandings from life-course research. Dev Neuropsychol. 2003;24(2–3):719–44.

247. Power C, Hertzman C. Social and biological pathways linking early life and adult disease. Br Med Bull. 1997;53(1):210–21.

248. Hallqvist J, Lynch J, Bartley M, Lang T, Blane D. Can we disentangle life course processes of accumulation, critical period and social mobility? An analysis of disadvantaged socio-economic positions and myocardial infarction in the Stockholm Heart Epidemiology Program. Soc Sci Med. 2004;58(8):1555–62.

249. Freedland KE, Miller GE, Sheps DS. The Great Debate, revisited. Psychosom Med. 2006;68(2):179–84.

250. Lundberg GD. Resolved: psychosocial interventions can improve clinical outcomes in organic disease—discussant comments. Psychosom Med. 2002;64(4):568–70.

251. Markovitz JH. Resolved: psychosocial interventions can improve clinical outcomes in organic disease—moderator introduction. Psychosom Med. 2002;64(4):549–51.

252. Relman AS, Angell M. Resolved: psychosocial interventions can improve clinical outcomes in organic disease (con). Psychosom Med. 2002;64(4):558–63.

253. Williams RB, Schneiderman N. Resolved: psychosocial interventions can improve clinical outcomes in organic disease (pro). Psychosom Med. 2002;64(4):552–7.

254. Oakes JM. Invited commentary: paths and pathologies of social epidemiology. Am J Epidemiol. 2013;178(6):850–1.

255. Taylor SE, Falke RL, Mazel RM, Hilsberg BL. Sources of satisfaction and dissatisfaction among members of cancer support groups. In: Gottlieb BH, editor. Marshaling social support: formats, processes, and effects. Newbury Park, CA: Sage Publications; 1988. pp. 187–208.

256. Prochaska JO, DiClemente CC, Velicer WF, Rossi JS. Standardized, individualized, interactive, and personalized self-help programs for smoking cessation. Health Psychol. 1993;12(5):399–405.

257. Mulder CL, Emmelkamp PM, Antoni MH, Mulder JW, Sandfort TG, de Vries MJ. Cognitive-behavioral and experiential group psychotherapy for HIV-infected homosexual men: a comparative study. Psychosom Med. 1994;56(5):423–31.

258. Amori G, Lenox RH. Do volunteer subjects bias clinical trials? J Clin Psychopharmacol. 1989;9(5):321–7.

259. Edlund MJ, Craig TJ, Richardson MA. Informed consent as a form of volunteer bias. Am J Psychiatry. 1985;142(5):624–7.

260. Gustavsson JP, Asberg M, Schalling D. The healthy control subject in psychiatric research: impulsiveness and volunteer bias. Acta Psychiatr Scand. 1997;96(5):325–8.

261. Holden G, Rosenberg G, Barker K, Tuhrim S, Brenner B. The recruitment of research participants: a review. Soc Work Health Care. 1993;19(2):1–44.

262. Leventhal T, Brooks-Gunn J. Moving to opportunity: an experimental study of neighborhood effects on mental health. Am J Public Health. 2003;93(9):1576–82.

263. Martinson BC, Crain AL, Sherwood NE, Hayes MG, Pronk NP, O'Connor PJ. Population reach and recruitment bias in a maintenance RCT in physically active older adults. J Phys Act Health. 2010;7(1):127–35.

264. Montgomery RJV, Borgatta EF. Family Support Project: final report to the Administration on Aging. Seattle: University of Washington, Institute on Aging/Long-Term Care Center; 1985.

265. Schneiderman N, Saab PG, Catellier DJ, Powell LH, DeBusk RF, Williams RB, et al. Psychosocial treatment within sex by ethnicity subgroups in the Enhancing Recovery in Coronary Heart Disease clinical trial. Psychosom Med. US: Lippincott Williams & Wilkins; 2004. pp. 475–83.

266. Glass TA, Berkman LF. The families in recovery from stroke trial (FIRST): a psychosocial intervention in stroke. Psychosom Med. 2000;62(1):1492.

267. Gibaldi M, Sullivan S. Intention-to-treat analysis in randomized trials: who gets counted? J Clin Pharmacol. 1997;37(8):667–72.

268. Hogan JW, Laird NM. Intention-to-treat analyses for incomplete repeated measures data. Biometrics. 1996;52(3):1002–17.

269. Newell DJ. Intention-to-treat analysis: implications for quantitative and qualitative research [see comments]. Int J Epidemiol. 1992;21(5):837–41.

270. Mohide EA, Pringle DM, Streiner DL, Gilbert JR, Muir G, Tew M. A randomized trial of family caregiver support in the home management of dementia. J Am Geriatr Soc. 1990;38(4):446–54.

271. The Multiple Risk Factor Intervention Trial Group. Multiple Risk Factor Intervention Trial: risk factor changes and mortality results. JAMA. 1982;248:1465–77.

272. Macintyre S. Good intentions and received wisdom are not good enough: the need for controlled trials in public health. J Epidemiol Community Health. 2011;65(7):564–7.

273. Ackermann RT, Holmes AM, Saha C. Designing a natural experiment to evaluate a national health care-community partnership to prevent type 2 diabetes. Prev Chronic Dis. 2013;10:E12.

274. Cerda M, Morenoff JD, Hansen BB, Tessari Hicks KJ, Duque LF, Restrepo A, et al. Reducing violence by transforming neighborhoods: a natural experiment in Medellin, Colombia. Am J Epidemiol. 2012;175(10):1045–53.

275. Diez Roux AV. Next steps in understanding the multilevel determinants of health. J Epidemiol Community Health. 2008;62(11):957–9.

276. Galea S, Link BG. Six paths for the future of social epidemiology. Am J Epidemiol. 2013;178(6):843–9.

277. Humphreys DK, Eisner MP. Do flexible alcohol trading hours reduce violence? A theory-based natural experiment in alcohol policy. Soc Sci Med. 2013.

278. Kaufman JS, Kaufman S, Poole C. Causal inference from randomized trials in social epidemiology. Soc Sci Med. 2003;57(12):2397–409.

279. Pinkston EM, Linsk NL, Young RN. Home-based behavioral family treatment of the impaired elderly. Behav Ther. 1988;19(3):331–44.

280. McCormick J, Skrabanek P. Coronary heart disease is not preventable by population interventions. Lancet. 1988;2(8615):839–41.

281. Galea S. An argument for a consequentialist epidemiology. Am J Epidemiol. 2013;178(8):1185–91.

282. Anonymous. Population health looking upstream. Lancet. 1994;343(8895):429–30.

283. Kaplan GA. Where do shared pathways lead? Some reflections on a research agenda. Psychosom Med. 1995;57(3):208–12.

284. McKinlay JB. The new public health approach to improving physical activity and autonomy in older populations. In: Heikkinen E, Kuusinen J, Ruoppila I, editors. Preparation for aging. New York: Plenum Press; 1995. pp. 87–103.

285. Rose GA. The strategy of preventative medicine. Oxford: Oxford University Press; 1992.

286. Frieden TR. A framework for public health action: the health impact pyramid. Am J Public Health. 2010;100(4):590–5.

287. Terris M. Epidemiology as a guide to health policy. Annu Rev Public Health. 1980;1:323–44.

288. Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, et al. Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. J Epidemiol Community Health. 2012;66(12):1182–6.

289. DiClemente RJ, Wingood GM. A randomized controlled trial of an HIV sexual risk-reduction intervention for young African-American women. JAMA. 1995;274(16):1271–6.

290. Kelly JA, St. Lawrence JS, Diaz YE, Stevenson LY, Hauth AC, Brasfield TL, et al. HIV risk behavior reduction following intervention with key opinion leaders of population: an experimental analysis. Am J Public Health. 1991;81(2):168–71.

291. Havas S, Heimendinger J, Damron D, Nicklas TA, Cowan A, Beresford SA, et al. 5 a Day for better health—nine community research projects to increase fruit and vegetable consumption. Public Health Rep. 1995;110(1):68–79.

292. Kristal AR, Patterson RE, Glanz K, Heimendinger J, Hebert JR, Feng Z, et al. Psychosocial correlates of healthful diets: baseline results from the Working Well Study. Prev Med. 1995;24(3):221–8.

293. Minkler M. Building supportive ties and sense of community among the inner-city elderly: the Tenderloin Senior Outreach Project. Health Educ Q. 1985;12(4):303–14.

294. Arnaud N, Broning S, Drechsel M, Thomasius R, Baldus C. Web-based screening and brief intervention for poly-drug use among teenagers: study protocol of a multicentre two-arm randomized controlled trial. BMC Public Health. 2012;12:826.

295. Campbell AN, Miele GM, Nunes EV, McCrimmon S, Ghitza UE. Web-based, psychosocial treatment for substance use disorders in community treatment settings. Psychol Serv. 2012;9(2):212–4.

296. Schulz DN, Kremers SP, de Vries H. Are the stages of change relevant for the development and implementation of a web-based tailored alcohol intervention? A cross-sectional study. BMC Public Health. 2012;12:360.

297. Strecher VJ, McClure JB, Alexander GL, Chakraborty B, Nair VN, Konkel JM, et al. Web-based smoking-cessation programs: results of a randomized trial. Am J Prev Med. 2008;34(5):373–81.

298. Dunton GF, Robertson TP. A tailored Internet-plus-email intervention for increasing physical activity among ethnically-diverse women. Prev Med. 2008;47(6):605–11.

299. Knowlden A, Sharma M. A feasibility and efficacy randomized controlled trial of an online preventative program for childhood obesity: protocol for the EMPOWER Intervention. JMIR Res Protoc. 2012;1(1):e5.

300. Kuijpers W, Groen WG, Aaronson NK, van Harten WH. A systematic review of web-based interventions for patient empowerment and physical activity in chronic diseases: relevance for cancer survivors. J Med Internet Res. 2013;15(2):e37.

301. Bond GE, Burr R, Wolf FM, Price M, McCurry SM, Teri L. The effects of a web-based intervention on the physical outcomes associated with diabetes among adults age 60 and older: a randomized trial. Diabetes Technol Ther. 2007;9(1):52–9.

302. Davis S, Abidi SS, Cox J. Personalized cardiovascular risk management linking SCORE and behaviour change to Web-based education. Stud Health Technol Inform. 2006;124:235–40.

303. Zheng K, Newman MW, Veinot TC, Hanratty M, Kim H, Meadowbrooke C, et al. Using online peer-mentoring to empower young adults with end-stage renal disease: a feasibility study. AMIA Annu Symp Proc. 2010;2010:942–6.

304. Lieberman MA, Golant M, Giese-Davis J, Winzlenberg A, Benjamin H, Humphreys K, et al. Electronic support groups for breast carcinoma: a clinical trial of effectiveness. Cancer. 2003;97(4):920–5.

305. Coulehan MB, Rossie KM, Ross AJ. Developing a novel Internet-based psychoeducational intervention for dementia caregivers. AMIA Annu Symp Proc. 2008:915.

306. Wu YH, Faucounau V, de Rotrou J, Riguet M, Rigaud AS. [Information and communication technology interventions supporting carers of people with Alzheimer's disease: a literature review]. Psychol Neuropsychiatr Vieil. 2009;7(3):185–92.

307. Centola D. An experimental study of homophily in the adoption of health behavior. Science. 2011; 334(6060):1269–72.

308. Centola D. The spread of behavior in an online social network experiment. Science. 2010;329(5996):1194–7.

309. Kuhn TS. The structure of scientific revolutions. 1st ed. Chicago, IL: University of Chicago Press; 1962.

CHAPTER 12

POLICIES AS TOOLS FOR RESEARCH AND TRANSLATION IN SOCIAL EPIDEMIOLOGY

M. Maria Glymour

INTRODUCTION

Medical policy is only a small component of *health* policy. Government rules and regulations affect nearly every exposure within the purview of social epidemiology.[1] Policies thus provide key avenues for translation[2] of evidence about the social determinants of health into population health improvements. In addition to this translation goal, policy evaluations can also address causal questions about the health consequences of the social factors shaped by the policy (a basic research goal). The translation goal answers questions such as "Does the Head Start program, which increases pre-school access for 4-year-olds, improve long-term health outcomes of those children?" The research goal answers questions such as "Does attending pre-school at age 4 improve long-term health outcomes of children?" It is worth distinguishing between these two goals, because there may be many other types of policies besides Head Start that improve pre-school access for 4-year-olds; thus the *research* question is more general and can help guide development of novel policies and programs beyond Head Start.

1 Governments are not the only institutions that set "policy," and the institutional rules of many other organizations—employers, manufacturers, retail organizations, schools, churches, etc. also have profound influences. Such institutional policies are equally relevant for both research and translation goals, although they have received even less research attention in social epidemiology.

2 We use the term "translation" to refer broadly to moving evidence from basic science research to population health improvements, e.g., via development of evidence-based policies, projects, or programs.

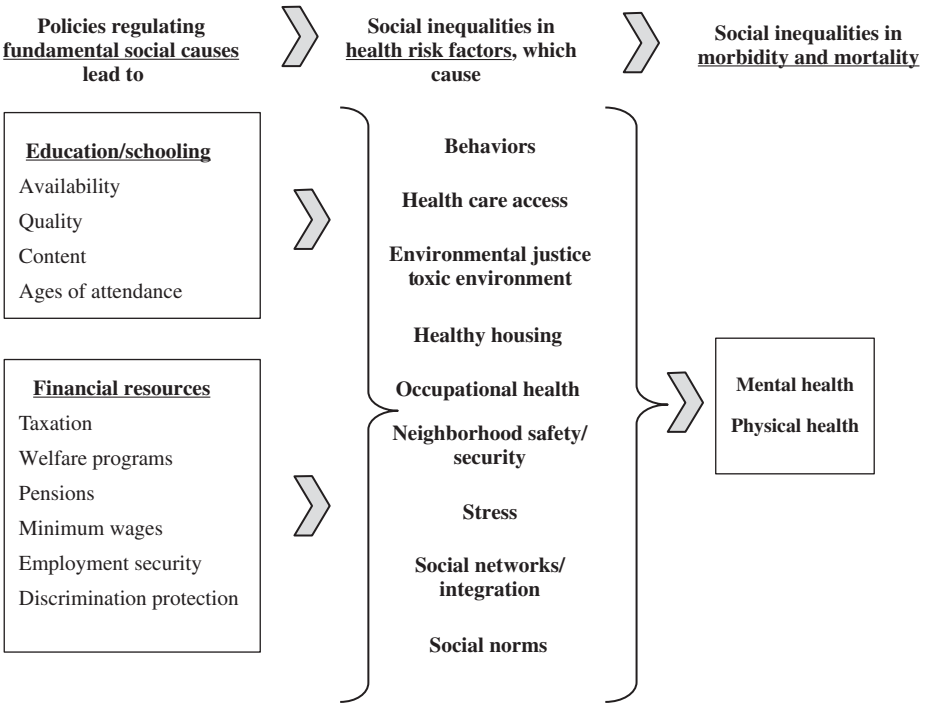


FIGURE 12.1: Policy mechanisms influencing social determinants of health and socially patterned risk factors. Some social policies directly influence "fundamental social determinants of health," such as education and income, but many social policies also target inequalities in health risk factors. Policies that eliminate social inequalities in health risk factors, for example by guaranteeing healthy housing in safe neighborhoods, can also reduce social inequalities in health.

The reach of social policies into the personal fabric of life can be surprising: legal standards influence how long we attend school and who our classmates are; our risk of living in poverty or falling into poverty if we become sick or unemployed; whether, when, and whom we marry; our rights to birth control and the number of children we have; and our religious practices. Policies also influence major pathways or channels linking the socioeconomic and psychosocial environment to health, such as toxic environmental conditions, health behaviors, or medical services (examples illustrated in Figure 12.1). Because of their wide-ranging and strong effects, social policies—including legal and regulatory decisions—should be a primary target for potential interventions to improve population health and reduce social inequalities in health.

A social determinants of health model implies that many policies that do not *primarily* target health, for example regulating employment, transportation, or housing, may have large unintended consequences for health. Unfortunately, important gaps in knowledge exist about the long-term health effects of most major public policies shaping socioeconomic conditions. Furthermore, social epidemiologists have often regarded policy evaluation as tangential to their primary research, rather than as the central task for understanding the social determinants of health.

WHY POLICY CHANGES ARE IMPORTANT TO HELP ANSWER CAUSAL QUESTIONS

With respect to the research goal of policy studies, changes in policies often create "natural experiments" to evaluate whether putative risk factors causally affect health and to illuminate how social determinants of health operate. Policy evaluations will not only inform us about the causal impact of interventions, but can also test more general causal hypotheses and guide theoretical development. The difficulty of identifying the causal effects of social resources is a theme throughout this book. Recent examples of discrepancies between results from observational studies and randomized trials or quasi-experimental studies suggest that observational evidence may in fact offer an unreliable foundation for designing social policies to benefit health (1–5). In this chapter, we emphasize studies that have used designs we think least vulnerable to confounding or reverse causation, i.e., those based on experimental or quasi- or natural-experimental data sources. Observational evidence may diverge from experimental evidence for many reasons, not all of which reflect faults of the observational study designs, but discrepancies between observational and experimental evidence clearly point towards the dangers of relying on observational evidence to understand the complex social mechanisms influencing health.

Social interventions are not without risks, and well-intended social policies can harm the intended beneficiaries under some circumstances. Even if the social intervention is not harmful, but merely ineffective, there is an opportunity cost when the money spent could be allocated to better uses. We need stronger evidence on causality and a more precise theoretical understanding of how specific, *modifiable* aspects of social conditions influence health. We conceptualize a tripartite process of evidence-based public health. This proceeds first from a theoretical understanding of the causes of disease, then evaluates the population health importance of specific causes, and finally evaluates the feasibility and effectiveness of specific interventions to eliminate causes of poor health (or, conversely, to create a health-promoting environment) (Figure 12.2). The model is iterative, as findings about causality based on interventions may shift theoretical understanding. Within this framework, evaluation of social policies plays an important role both in establishing the causes of disease (step 1) and as a tool for intervening on major determinants of health (step 3). It provides a real opportunity to impact population health rather than to just identify risks.

Although this is changing rapidly, much of the prior research on long-term effects of social policies has been conducted in economics or public policy departments, with relatively little participation of epidemiologists. Health is often poorly measured in policy studies; heterogeneous effects on different health outcomes are not considered; and neither hypotheses nor interpretations are closely linked to understanding of the social or physiologic pathways to disease. We hope that this chapter will serve as an impetus for epidemiologists to consider the health consequences of social policies as a core domain for epidemiologic research. We also hope that it may cultivate increased interdisciplinary engagement across policy-based fields by highlighting some of the special considerations relevant to studying health.

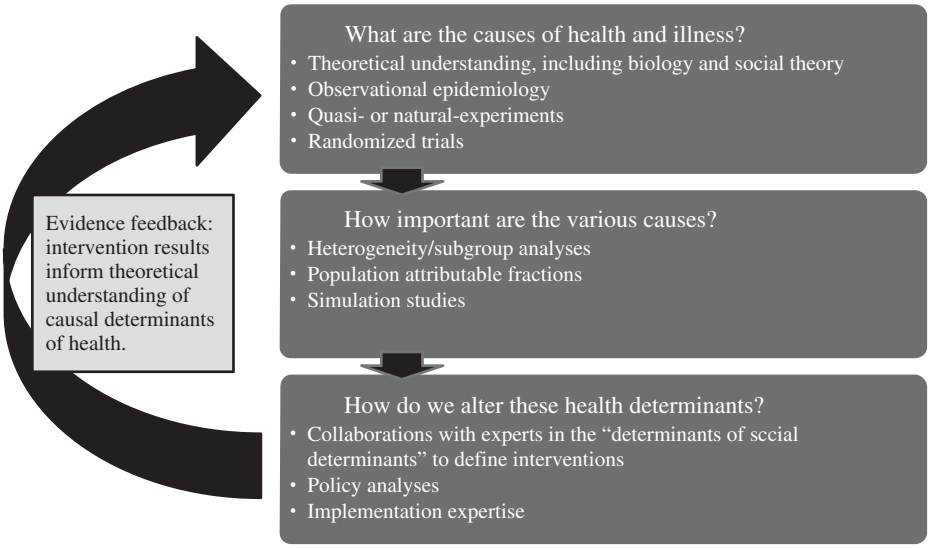


FIGURE 12.2: Steps in the development of evidence-based public health.

COMPARING POPULATION HEALTH BENEFITS OF UPSTREAM VERSUS DOWNSTREAM STRATEGIES

"Upstream" social factors influence health via a cascading sequence of "downstream" events and exposures. Upstream determinants such as education and income are often called "fundamental" causes of health because they are flexible resources that can be deployed to improve a wide array of health outcomes in extremely diverse settings (see Chapter 2). Empirical evidence reveals reinforcing feedback processes between health and socioeconomic status (SES), so it is an oversimplification to dichotomize upstream (fundamental) versus downstream (proximal) factors. But this simplification is useful to distinguish between policies that influence social resources very broadly—leaving to the individual decisions about whether and how to use those resources to improve health—and policies that attempt to deliver specific health-promoting resources.

For example, many policies deliver or restrict access to education and income. Individuals in turn deploy their education and income to avoid more proximal health risk factors or to gain access to more proximal health-promoting resources, such as healthy housing or medical care. On the other hand, another large set of social policies do not target SES but affect these downstream mechanisms directly. In Figure 12.1, we represent this flow, from policies affecting "fundamental" causes on the left, to policies affecting mediating health risk factors that are socially patterned, and finally culminating in health.

Considering Figure 12.1, we can contrast a strategy to address socioeconomic inequalities in health via policies influencing variables at the far left, e.g., education and poverty, to a strategy that targets factors further right, e.g., workplace safety. Either approach might achieve

substantial improvements in public health. Both approaches should be considered important tools in the repertoire of strategies to redress social inequalities in health.

Is there any reason for a general preference for "upstream" versus "downstream" policies, for example based on the population health impact that might be achieved via these alternatives? There are advantages to upstream interventions that change the distribution of social resources. Fundamental cause theory suggests that policies that operate upstream on the distribution of resources may have large impacts because social disadvantage is like a hydra's head: interrupting any single mechanism between a fundamental cause and health is futile because it will only be replaced by another pathway. Further, efforts to interrupt downstream consequences of social determinants of health have greater potential for undesirable long-term consequences because they may distort incentives or inappropriately siphon resources from one domain to another. For example, maternity leave policies are sometimes structured so that employers must bear the cost of leave. Such policies create a financial incentive for employers to discriminate against women of child-bearing age. We can regulate against gender discrimination, but such laws may be difficult to enforce. Thus, it may simply be preferable to design policies such that financial incentives align with respecting gender equality. Tax subsidies for employer-provided health insurance may lead to overuse of expensive medical procedures, even if many recipients of the excess care would prefer to spend the extra money for something other than medical care. These are all arguments about possible risks of downstream regulation, and they would support upstream interventions. Addressing "upstream" inequalities may have other highly desirable social benefits besides health; we may desire to live in a society with more education and less poverty, regardless of the health implications.

Despite the theoretical advantages of upstream interventions, it may often be more politically feasible to interrupt mediating pathways than to redistribute upstream resources such as income. Further, policies that target more proximal mechanisms rather than distal social determinants often have somewhat stronger evidence supporting their influence on health (precisely because it has been possible in the past to achieve changes in such policies, so that we can learn about the consequences of those changes). The "flexible resources" delivered by policies targeting upstream factors such as income may well be used for non-health related activities, or even for activities that are detrimental to health. Finally, some degree of socioeconomic inequality might be nearly inevitable (although "inevitability" is called into question by empirical evidence showing that the magnitude of these inequalities varies tremendously over time and place (6–8)).

Cross-national comparisons can provide some evidence on the potential health impact of upstream strategies targeted at reducing inequalities in educational attainment or income versus downstream strategies targeted at reducing the health consequences of such inequalities. Older Americans average higher education levels than older Europeans, but the difference in health between individuals with high versus low education is larger in the US than in Europe. One can thus roughly conceptualize the US as representing the health outcomes achieved by a country with an *upstream* strategy (increasing average levels of education but tolerating substantial differences in health between those with high versus low levels of education) and European countries as pursuing downstream strategies (closing the gap in health between individuals with high versus low levels of education).[3] Recent

3 This is of course not a perfect correspondence with actual policy intent—many US policies are intended to reduce inequalities between people with high versus low socioeconomic status and many European policies are intended to increase average educational attainment.

work contrasts the hypothetical life expectancy if the US had the same *educational distribution* as European countries, but the existing US level of health inequalities (i.e., much worse inequalities than Europe).[4] For example, during the 1990s, US women had higher mortality rates than French women by 419 deaths per 100,000 person-years. In both countries, education is inversely associated with mortality, but compared to French women, US women have much higher average education levels. If US women had the same education as French women, the US mortality disadvantage would be expected to deteriorate by 192/100,000, a 45% increase in the US-France gap (again, assuming education is causal). Alternatively, we could imagine a world in which the US had the existing educational distribution, but a level of inequality between educational categories similar to that prevailing in France; in this situation, the US disadvantage would be reduced by 183/100,000, a 44% reduction in the gap (9).

If the link between education and health is causal, it suggests that the US investments in education have substantially increased life expectancy compared to Europe, but that the huge health inequality between low- and high-education women in the US has offset this benefit. These results suggest that either an upstream or a downstream strategy—increasing education levels or reducing health inequalities between education levels—could plausibly have similar population health benefits. We do not know how general this result is, but it is helpful to recognize that social inequalities in health can potentially be addressed by targeting the so-called fundamental cause, or, alternatively, by targeting the mediators.

EVIDENCE ON UPSTREAM POLICIES INFLUENCING SOCIOECONOMIC RESOURCES

In previous chapters, we reviewed evidence on the health effects of education, income, and other "fundamental" causes. Much of this research on the health effects of SES was drawn from policy-based research, such as compulsory schooling law changes. Rather than review that research here, we mention a few studies that focus more specifically on the effects of the *policies*, rather than using the policies to estimate effects of SES, and we highlight a few distinctions in the types of policies regulating SES. Of course, education and income are directly influenced by a host of policies, but the effects of specific policies on individuals are often quite small. When considering the public health importance of policies, note that even if the policies have very small effects on individuals, they may have very large effects on the population as a whole. A small benefit delivered to everyone may have a bigger population impact than a large benefit for just a few people.

Social policies relevant to education influence both educational attainment and educational quality. Mandatory schooling laws, regulating age of school enrollment and legal school dropout

4 This thinking also sets aside issues of causality and simply says, if we assume these policy frameworks are entirely causal, how much could plausibly be achieved by changing the distribution of education in Europe versus changing inequalities between educational groups in the US?

or work ages, have been shown to modestly increase years of completed schooling, although such laws have historically been selectively enforced. Because compulsory schooling laws were recorded in a standardized fashion and changed both in the United States and internationally during the 20th century, they have been used as the basis for several studies of the effects of education on health (see Chapter 2). Examining the long term health effects of changes compulsory schooling laws in low and middle income countries is an especially important next step in this research area, because the effects of education may differ in communities with lower average education levels. Many other policies likely influence education as much or more than compulsory schooling laws, but have received very little attention in health research. These include policies regulating funding for primary, secondary, and tertiary education; subsidies for education outside the public school system; curriculum standards; court decisions on desegregation (10); availability of kindergarten and other pre-primary educational experiences; and adult education initiatives, such as general educational development (GED) credentials or employee training initiatives (11).

Myriad policies target income via taxation, unemployment insurance, financial subsidies to poor families, disability insurance, and pension systems. These policies succeed in their primary goal: they increase the financial resources of beneficiaries and reduce inequalities in income. In 2004, nearly 16% of US families were shifted above the poverty line because of social programs such as old age Social Security, the Earned Income Tax Credit, and Temporary Assistance to Needy Families. In other words, if such programs had not existed, 29% of US families would have been below the poverty line; after considering benefits from these programs, only 13.5% of US families were below the poverty line (12).

But do these policies succeed in improving the health of beneficiaries? If so, do all policies succeed equally well, or are some especially effective? One might argue that policies benefiting the elderly would have the largest health impacts because the elderly have the most health needs. On the other hand, lifecourse models with "sensitive periods" in childhood would imply that investments in children would have the largest health impacts. Such contrasts are largely theoretical speculation: there is not much empirical evidence and it is far too scanty to support careful comparisons of the likely health impacts of investments. We can identify many other policy features that might be important to shape the health effects of the policy. For example, policies might redistribute money between individuals (typical taxation policies); or might primarily smooth financial resources across an individual's lifecourse so he or she does not have to endure sudden declines in standard of living due to illness, unemployment, or old age. Many policies may achieve both types of redistribution (e.g., certain pension plans). The health effects of increasing long-term average income may differ from the health effects of reducing short-term fluctuations in income or smoothing over periods of special vulnerability. Smoothing schemes may influence health both via the immediate consequences of insufficient financial resources and via anxiety or mental health consequences of worrying about financial insecurity, even in relatively good times.

Pension policies, such as Social Security, may have especially large population health effects because they influence income for many individuals and often involve large transfers. Even here, however, the evidence is limited to a handful of quasi-experimental studies. Social Security retirement benefits dramatically reduced poverty rates among elderly Americans (13), but to date there have been no compelling evaluations of the health effects of this policy. Snyder et al attempted to estimate the mortality effects of Social Security income based on the "notch" generation, who received extra income due to an administrative correction in Social Security benefits. After-65

mortality was unexpectedly *elevated* among men born in the higher-income cohorts (the second half of 1916) in comparison to cohorts receiving lower Social Security benefits (born in the first half of 1917) (14). However, this research has been criticized because the "notch" birth cohort mortality differed from mortality in adjacent birth cohorts even prior to retirement (15). More recent work compared trends in mortality with the introduction and expansion of Social Security, but could not account for other sources of secular trends in health (16).

In a more convincing analysis, Herd et al. took advantage of year-to-year variations in state Supplemental Security Income (SSI) benefits to show that higher benefits predicted lower disability rates (mobility limitations as reported on the US Census) (17). Effects of the policy *per se* were moderate—a $1200 increase (in year 2000 dollars) in the maximum annual SSI benefit predicted a 0.46-percentage point decline in the probability of having a mobility limitation among unmarried elderly and a 1.8-percentage point decline among those unmarried elderly who were in the lowest quartile of income. Only a small fraction of unmarried elderly received SSI, however, so the impact on individuals who actually received SSI was estimated to be nearly a 20-percentage point decline in probability of disability. This would correspond to an estimated cost of around $7000 to prevent one disability case. If causal, this benefit would likely be largely paid for by reductions in medical costs. For example, Medicare spending on individuals with a limitation in an activity of daily living is approximately $4000 to $9000 per year (depending on the type of limitation) greater than spending among individuals with no limitations (18).

Additional work has examined the effects of family income support policies on children's health. Evidence on the effects of Aid to Families with Dependent Children on birthweight is inconclusive, with some studies indicating benefits and others indicating harm, depending on the estimation method. Effect estimates from the most rigorous analytic approaches are quite imprecise (19). The Earned Income Tax Credit (EITC) has been shown to increase body mass index and obesity among women, with approximately a 3-percentage point increase in probability of obesity per $1,000 annual benefit (20). Evidence on the effects of EITC on smoking is mixed, with some studies reporting increases in smoking probability (21), but several other studies reporting the contrary (1, 22, 23). The only study to assess the effects of EITC and birthweight suggests that the policy improves birthweight (1).

As illustrated with EITC, the health benefits of income support policies cannot be taken for granted. Health consequences may differ across health outcomes and depend on how the additional income is delivered (see also Chapter 2). Annual payments to Cherokee nation tribe members are based on profits from casinos built on tribal lands, and dispersed in lump sum payments at scheduled dates. A combination of qualitative and quantitative evidence indicates that these payouts are associated with spikes in accidental deaths (24). Other research using the introduction of these payments suggests beneficial effects on child psychopathology (25, 26). Similarly, a randomized trial of Connecticut welfare reform showed that the experimental employment-mandate policy applied to welfare recipients succeeded in increasing employment and earnings, but had no apparent effect on mortality (27). One explanation for the lack of mortality reduction was that gains due to improved earnings were offset by occupational hazards and increased stress associated with the job search.

The strongest evidence for the health benefits of income support policies comes from international contexts in which the additional income is likely to be a large percent increase in household income. For example, Oportunidades randomly assigned communities to implement a

conditional cash transfer (CCT) program; implementation of the benefits in control communities was delayed by 18 months to allow evaluation of the health effects of the program. Conditional cash transfers improved physical and cognitive health outcomes for children whose families benefited (28, 29). Recent findings suggest there may be important spillover benefits of CCT programs for elderly members of the household, even though CCT programs are not targeted to older adults (30). Oportunidades improved the health of the older population for a number of indicators, with larger effects for recipients with a longer period of time receiving benefits. Most of these health effects were concentrated among women.

A CCT program in New York City was designed based on the success of Oportunidades, but reported mixed results. There were few overall improvements in academic achievement, but there were improvements in self-rated health and in academic outcomes for the subgroup of high school children who were "well-prepared" in 9th grade (31). The lesson from the discrepancy between the success of CCTs in low-income rural Mexican families versus modest outcomes in low-income urban US families may be that CCTs must be designed with a detailed understanding of the barriers to success within the context of the individuals being served (28, 29). As such, they must ensure that both the conditions and the cash transfer in the conditional cash transfer are appropriate to address those barriers. Money may simply not be enough to enable people to achieve complicated, difficult goals, such as performing well in school or losing weight.

Particularly compelling evidence of the effects of income redistribution policies emerged from post-apartheid changes in the pension system in South Africa. After the dissolution of apartheid, black South Africans were included in the pension system and older blacks received unexpected financial support. These older individuals often lived with large families and pooled income within the household. In families that pooled income, living with pensioners predicted better health for all family members. For example, indicators of healthy growth for children improved in families living with women eligible for a pension. In other words, grandmothers received extra money, and this income benefited their granddaughters who were living with them (32, 33).

Overall, the evidence that income support policies promote better health—particularly children's health—is promising, but the evidence is fairly scant. In part this represents the intrinsic methodological challenges in deriving valid effect estimates. The most persuasive study designs are based on policy changes and, in particular, changes implemented at different times in different places. Such changes often, though not always, induce relatively small differences in income compared to the overall variation in family income.

Further, the most rigorous designs do not compare individuals who receive the benefits to individuals who do not receive benefits, because individuals who receive benefits are likely to differ from non-recipients on countless unobserved characteristics that also influence health. Rather, analyses identify something that "pseudo-randomizes" individuals in the population to receive the extra income, e.g., measurable characteristics that change their probability of being eligible, or place-level variations in program generosity. For example, in a study of the effects of unemployment benefit generosity on suicide risk, Cylus et al., examined differences in suicide risk between states over time (34). Unemployment benefits are presumably most relevant for the unemployed and their families, but comparing the employed to the unemployed introduces confounding by determinants of employment. Instead, Cylus used state average suicide rates (stratified by demographic characteristics such as age and sex), so any effect of the unemployment benefits was estimated based on pooling a small number of unemployed beneficiaries with a much larger number

of people who were employed. Adopting this strategy (rather than comparing suicide rates among people receiving generous benefits versus people receiving limited benefits) avoids possible reverse causation or confounding, but in this type of analysis, effect estimates are likely diluted compared to effects on people who actually receive the benefit. An additional challenge in such a design is that the etiologic period linking increased income to health outcomes is of unknown duration. To detect health effects of such small differences, large samples are necessary. Overcoming these challenges requires large longitudinal data sets or surveillance data linking information on place of residence (and therefore policy environment) to health outcomes. Lessons from the research to date should guide future research in the following ways: *(1)* health should be assessed along multiple dimensions and at multiple points in the lifecourse; *(2)* beneficiaries may include family members or other network members as well as the primary individual; *(3)* key mechanisms may include both material and psychological consequences of the additional income; and *(4)* specific features of the delivery of extra income may be influential.

POLICIES TARGETING SOCIALLY PATTERNED HEALTH RISK FACTORS

What about evidence on the effects of policies that target downstream, mediating pathways? This umbrella—policies that affect mediating variables between SES and health—covers countless policies. Because the effects of social conditions are so far-reaching, nearly all epidemiologic research addresses risk factors that may mediate social inequalities. For example, nutritional, clinical, and environmental epidemiology are all relevant, because diet, medical care, and environmental toxins are all affected by SES.[5] All of these factors are, of course, influenced by various policies. Here we briefly mention a set of mediators thought to be of special relevance for social inequalities in health, and the types of questions in these areas that speak to social inequalities. This list is meant to be illustrative rather than comprehensive.

ENVIRONMENTAL TOXINS

Toxic exposures are an important mediator between social disadvantage and poor health: exposure to dangerous and noxious environments is much more common among poor families. Strong evidence supports the health impact of environmental policies. This evidence arises from both meticulously conducted observational studies and quasi-experimental studies evaluating factors such as air quality (35) and lead in gasoline and paint (36, 37). Just as with cigarette debates, legal regulations of lead were supported by extensive scientific evidence and achieved only via bitter

5 Quantifying how social conditions influence exposure to such mediating risk factors (e.g., nutrition, medical access) can help identify opportunities to reduce social inequalities in health. Such evaluations have an additional benefit because social factors probably confound the estimated effects of these mediating risk factors on various health outcomes. In other words, understanding the social determinants of proximal risk factors is important in order to derive causal effect estimates for the proximal risk factors.

political disputes. Herbert Needleman, who conducted some of the decisive research demonstrating that even low levels of lead exposure harmed children's cognitive development, recounted the process of developing Environmental Protection Agency regulations of leaded gasoline (which was at that time routinely added as an "anti-knock" agent to gasoline). Needleman recalls his frustration after asking a Dupont scientist:

> " 'You have these PhDs, these smart chemical engineers, why don't you develop a better anti-knock agent [one without lead]?" and he said "Well, Herb, to tell you the truth, our economists are looking at the gasoline market. It's beginning to flatten out. There's not going to be the same kind of demand. And we're not going to put 100 million dollars into R and D." This is what he said. This was my post-postgraduate education. That all this b-s-ing in the criteria document [about the lack of any danger from lead in gasoline] didn't mean anything. Dupont's scientific position was determined by the company's economists." (38)

The vigor of the industry opposition reflected the financial costs of making the proposed changes, and regulation to reduce lead levels was only achieved because the scientific evidence for both the health and financial benefits was rigorous and overwhelming.

The benefits of improvements in many environmental exposures are wide-ranging and well documented, although debate remains for specific pollutants about "safe" exposure levels and the preferred approaches to regulations (e.g., cap and trade). The long-term effects of environmental exposures are likely compounded for children because air pollution and lead exposure influence academic achievement and educational attainment. Fundamental cause theory suggests that policies mandating healthy environments will likely influence not only overall population health *but also reduce health inequalities*. Individuals with high education and income can often avoid unhealthy environmental exposures, whereas disadvantaged people cannot.

Despite the substantial evidence base, important questions remain unanswered. Key topics for social epidemiologists working at the intersection of environmental epidemiology include additional evaluations of effects of specific environmental policies; estimating net effects of hypothetical policy changes on health, health inequalities, and health expenditures; and evaluating potential health effects of unintended consequences of environmental regulations, such as increased housing costs.

FOOD POLICY

Food policy has important consequences for social inequalities in health, and much food policy is specifically intended to protect poor families from food insecurity. Nutritional support policies raise an important challenge because support for food, like many resources, is fungible. In-kind resources can, at least partially, be translated into extra money and applied to other goods. For example, although food stamps nominally provide subsidies for food, each dollar of food stamp support is estimated to increase food expenditures by less than 50 cents (39). Despite this, nutritional supplement programs such as food stamps and the Supplemental Nutrition Assistance Program (SNAP) reduce food insecurity and also appear to reduce obesity in children (40), though

some evidence suggests increased obesity in adult female SNAP beneficiaries (41). Substantial debate relates to whether food supplement programs should restrict covered food to healthful items. Given the fungibility of food supplement money (i.e., all families must spend some money on food, so if a family receives an extra dollar earmarked for spending on food, it frees up a dollar to spend on other things), such restrictions might not have substantial benefits on actual food intake, but this remains a critical policy question.

MEDICAL CARE

Illness has major direct and indirect financial consequences, via medical care expenses and reduced participation in the paid labor force. The relative importance of meaningful access to affordable and effective medical care in creating health inequalities is uncertain, but it is most likely a contributor. Medical care may also be an important mediator of psychosocial determinants of health. For example, recent evidence suggests that elevations in mortality shortly after spousal bereavement may be partially, although not fully, attributable to health differences that emerge before bereavement, and specifically to deterioration in health care, including medication coverage (42, 43). Evidence on the intersection of social factors and medical care can guide redesign of medical systems and improve quality of care. For example, social predictors can be used to identify at-risk populations for primordial and primary prevention efforts; identify individuals for whom the existing care system is likely to fail; and show heterogeneity in treatment effects to guide "personalized" medical care (44) (Figure 12.3). Although only some people are chronically disadvantaged, nearly everyone experiences episodes of increased social vulnerability, such as bereavement or short-term spells of financial insecurity. Redesign of health care delivery systems to optimize quality of care for individuals who are chronically at risk will likely also improve quality of care for individuals who are not chronically disadvantaged but who experience transient periods of elevated social risk.

Given the burgeoning prevalence of major chronic conditions, such as obesity, hypertension, and diabetes, access to high-quality care to help manage these conditions is likely to become even more important in the future. For example, nearly a third of Americans aged 20 and over and more than half of Americans over age 55 have hypertension (45). Of course, the care necessary for people to control their hypertension may well include providing resources that are well beyond typical clinical services. Health insurance subsidies may improve health directly by providing medical care, and also improve health indirectly via economic benefits. Further, because many people recognize that they do not have financial resources to weather a major health event, lack of health insurance may be an ongoing source of stress and anxiety.

The evidence for potential health benefits of providing access to high-quality medical care and health insurance is fairly compelling, based on findings from the Oregon Medicaid experiment (46), quasi-experiments based on the 1980s Medicaid expansion (47), and analyses of the impact of Medicare creation in 1965 (48). An influential study documented marked declines in maternal mortality in southern states when the civil rights act eliminated racial segregation in hospitals. The grim implication is that black women had been dying because they were denied access to hospitals serving whites (49). The Oregon Medicaid experiment showed that access to Medicaid reduced out-of-pocket spending, catastrophic expenditures, medical debt, and

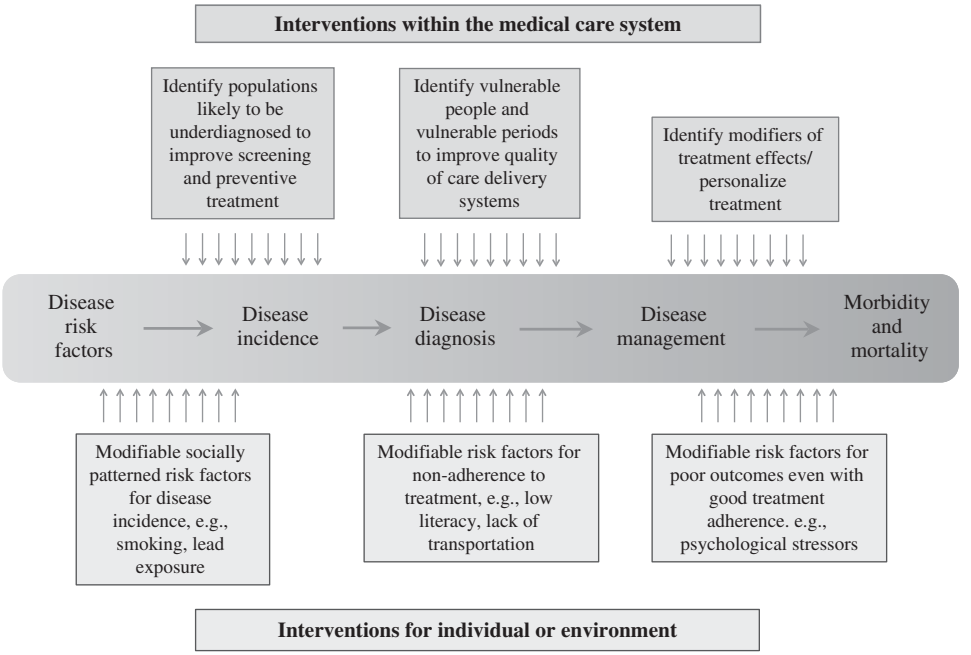


FIGURE 12.3: Conceptual model for strategies translating social epidemiology research into interventions to reduce morbidity and mortality: the Social Epidemiology Translation (SET) framework.

bill-skipping for participants, and significantly improved mental health (46). There was little evidence to demonstrate improvements in physical health in the short term, since the sample was too small to provide informative confidence intervals. However, rates of mammography, pap smears, diabetes diagnoses, and use of diabetes medication all increased. The group randomized to receive insurance had higher rates of diabetes diagnoses; this was not because insurance caused these individuals to develop diabetes, but rather they were diagnosed because they were insured (rates of *undiagnosed diabetes* were higher in the control group). This highlights the importance of choosing an appropriate outcome for evaluations of the health effects of policies. Mammography, pap smears, and diabetes diagnoses are generally thought to improve long-term outcomes, so—with a caveat acknowledging the various possible iatrogenic consequences of increased medical care—the anticipated long-term benefit of the health insurance looks promising.

RISKY HEALTH BEHAVIORS

Policies that influence behavior—either directly or indirectly, for example, via taxes—are recognized as crucial to health and health inequalities. Countless policies regulate behaviors relevant to health, such as diet, seat belts or bicycle helmet usage, firearm ownership, smoking, or medication adherence, sexual behavior and birth control (see also Chapter 10 and 13). These policies range widely from passive health education/information programs to initiatives to modify social norms or defaults to highly restrictive and specific mandates, such as smoking bans (see the "Label, nudge, or tax" debate: (50)).

For many of these policies, the evidence for health benefits is overwhelming. Public smoking bans have received particular attention. Meta-analyses estimate that enactment of smoke-free legislation is associated with 15–40% reductions in hospitalization rates for cardiovascular, cerebrovascular, and respiratory conditions, with largest benefit for younger individuals and non-smokers (51). This estimate may be sensitive to modeling assumptions regarding secular trends in cardiovascular disease, a common limitation in natural experiments (52). An important remaining question is how such policies affect social inequalities in health and spillover effects, or unintended consequences, of policy changes. The most restrictive behavioral regulations, such as seat belt mandates or smoking bans, may be the most effective strategies to narrow social inequalities in health if individuals from lower socioeconomic strata are least likely to adopt behavior change voluntarily. For example, seat belt usage is among the behaviors with the most marked improvements in both prevalence and inequalities (53, 54).

OTHER PROMISING AREAS

Social policies in many other domains regulate resources that may partially mediate social inequalities in health. Particularly active areas include healthy and affordable housing, residential segregation, discrimination, marriage and family law, unemployment, and immigration policies. A large single-blinded randomized trial of retrofitting housing with insulation in low-income communities in New Zealand reported significant benefits for self-rated health and reports of children taking fewer days away from school and adult household members taking fewer days away from work (55); other housing intervention studies corroborate the (perhaps unsurprising) finding that warm, dry housing improves health outcomes of residents (56). For many other types of social policies, the evidence for the causal effects of the targeted resource is mixed. For example, the Moving To Opportunity (MTO) trial randomly assigned families living in public housing developments to receive vouchers for private market rental subsidies. Some families were randomly assigned to vouchers that could only be used in low-poverty communities. Although MTO reported many encouraging results, such as reduced obesity and lower diabetes in mothers, effects on many outcomes were null (2). For some groups, particularly adolescent boys, findings indicated that moving to low-poverty neighborhoods was associated with *worse* outcomes, a result that remains largely unexplained (57).

MOVING FORWARD

The last decades have seen a flourishing in thinking about causal inference in many disciplines (58). These developments have helped to clarify the definitions of causal effects, how various statistical parameters correspond (or fail to correspond) to plausible consequences of interventions in the real world, and what types of data we need to collect to answer substantively important questions. Rigorous evidence on which exposures or conditions causally influence health is the first step in building an evidence base for public health, and providing such evidence is extremely challenging for many of the exposures of interest in social epidemiology. Many of the generally well-trusted

research methods in conventional epidemiology—based on longitudinal observations to establish a temporal sequence of exposure before outcome, accompanied by regression adjustment for a set of measured confounders—are less convincing in research on social determinants of health. Conventional observational analyses rest on the assumption that, after accounting for the set of measured confounders, exposure is effectively randomly assigned. With respect to many social exposures—education, income, occupation—this assumption is often implausible, even when a long list of measured covariates is available. Thus some of the most convincing evidence on causal effects of social determinants of health has emerged from alternative research designs, including approaches that are more commonplace in other disciplines.

QUASI-EXPERIMENTS

Quasi-experimental approaches to evaluating policy effects on social and socioeconomic determinants of health are flourishing, and the many new examples of policy analyses based on pseudo-randomization are extremely promising for health research. Instrumental Variables (IV) analyses of data generated from quasi-experimental or pseudo-randomizing events can be used to estimate the causal effect of the treatment on the outcome even if there are unobserved common causes of the exposure and the outcome. Recent studies have leveraged, for example, preferences and biases of pseudo-randomly assigned judges. In health research, this idea has given rise to "physician preference" IV analyses, in which the average prescribing pattern of the treating physician, rather than any individual patient characteristics, is used to evaluate effects of a particular prescription medication (59, 60). The validity of the physician preference IV is controversial because characteristics of the treating physician are likely correlated with characteristics of the patient. This challenge, however, could potentially be overcome with larger, more comprehensive, data sets that included many patients for each physician (thus allowing models that controlled for each physician using fixed effects). Many statistical techniques entail a tradeoff between precision/variance and validity/consistency such that the best designs to rule out bias also have the least statistical power. Increases in sample size may make such tools more useful in the near future. More credible examples in social policy include a study on the effects of parole versus incarceration on juvenile offenders on high school completion and recidivism. This study took advantage of the fact that juvenile court judges were effectively randomly assigned to defendants, and some judges were more lenient than others (61). A similar design used the fact that Social Security disability applications are decided by an arbitrarily assigned case reviewer to evaluate the effects of Social Security disability benefits on employment outcomes (62).

ANALYTIC TOOLS

Recent research in health has made more frequent use of a set of related analytic approaches, including fixed effects, difference-in-difference (or difference-in-difference-in-difference) specifications, regression discontinuity, and IV analyses (10, 63, 64). Difference-in-difference models, for example, estimate the effects of a policy on the prevalence of an outcome by contrasting: *(1)* the change in prevalence of the outcome before and after the implementation of the policy in a group

of people who were affected by the policy to *(2)* the difference in the prevalence of the outcome that occurred during the same period for a group of people who were not affected by the policy. For example, when a calorie-posting policy was implemented in 2008 in New York City chain restaurants, but not in Boston restaurants, researchers compared the change in caloric beverage purchases before versus after the new policy in New York to the change in caloric beverage purchases during the same period of time in Boston (where no policy was implemented) (65). The central idea is that although New York and Boston may differ in general, they would have experienced the same trend over the course of 2008 had it not been that New York introduced a requirement that chain restaurants post calorie information. The credibility of difference-in-difference approaches rests on theoretical understanding of the process in question and the plausibility of the negative control (i.e., *whose* differences are being contrasted) (66, 67). Negative controls are usually defined based on population subgroups that should not be affected by the policy in question but *should* be affected by other determinants of the outcome in question. This is most convincing when there is a plausible hypothesis about which subgroups should be affected by the policy, and when multiple comparison groups can be identified. For example, Liu estimated the effects of desegregation on pregnancy rates of black adolescents by contrasting changes in teen pregnancy rates between decennial censuses (the first difference) in places that did or did not experience school desegregation (the second difference) during the follow-up period. Liu supported inferences by using a difference-in-difference-in-difference analysis with comparisons to *both* similarly aged white females and slightly older black females as negative control groups. Her findings showed that changes in black teenage pregnancy rates were specific to places that desegregated and specific to black females in the relevant age groups (10).

Negative controls can also be based on *outcomes* that should not be affected by the policy in question but should be affected by other sources of bias, e.g., secular time or social factors related to policy benefit eligibility (66). Lipsitch et al. refer to this as the "check-for-an-effect-impossible-by-the-hypothesized-mechanism" approach and point to the example of evaluating influenza vaccines on hospitalization, with a negative control of hospitalization for trauma (which could not plausibly be affected by the influenza vaccine). Approaches that use negative control outcomes not to only assess the presence of confounding but also, in some situations, to actually "back out" or quantitatively account for bias due to unobserved confounders have recently been developed (68).

EXPANDING THE QUESTIONS

A final area of important innovation relates not to the methods of identifying policy effects, but to the range of questions being asked about policy effects. Interesting expansions include focusing on how the *timing* of policy exposure influences outcomes (26, 64). These analyses incorporate recognition of alternative lifecourse etiologic models (reviewed in Chapter 2) and will provide invaluable guidance for developing policies with the most beneficial health effects.

As we recognize potential adverse consequences of some policies, it is also important to expand our research questions to address unintended consequences. For example, it is very difficult to prevent certain types of discrimination via strictly legal regulations, because inter-personal discrimination is often difficult to document. Although we can document racial discrimination

in employment and housing via audit studies (in which "fake" job or housing applicants of different race/ethnic backgrounds but similar qualifications apply for the same employment or housing opportunity (69, 70)), identifying discriminatory decisions in realistic settings is often more difficult. Because it is difficult to document such discrimination, legal regulations against discrimination are often ineffective. Mandating benefits in order to protect some groups, e.g., maternity leave mandates that are financially costly to the employer, may even have adverse consequences for the intended beneficiaries. For example, the costs of such mandates may be shifted to the employee, or the employer may respond by increasing employment discrimination (e.g., refusing to hire women of child-bearing age); formal evaluations of the net impact of such policies are important (71).

Research on how policy benefits spill over to others in the family or to the network of the primary beneficiary is also of special importance for extensions in social epidemiology. As described in Chapter 7, relationships with others can influence health via myriad mechanisms, but this recognition has not been fully incorporated into evaluation of the health effects of social policies. Results from the South African pension experiment and from the Mexican Oportunidades program suggest direct resource sharing, and there is substantial evidence on the likely benefits of maternal education for children's health. Other potential spillovers could occur less directly. Policies may influence behavior modeling or norms; for example, if workplace smoking bans reduce employee smoking they may indirectly reduce the chances that the employees' children will start smoking. Harmful policies may also have domino effects: extremely high rates of incarceration for young African-American men may adversely affect the financial and social stability of their children.

The South African pension expansion and the Oportunidades studies also stand out because they tell us about effects in extremely poor communities. Understanding heterogeneity of effects is a critical next step, i.e., who benefits most and who benefits least from proposed policies? Effects may differ based on individual characteristics (e.g., a person's own level of poverty) or on community characteristics (e.g., national poverty rates, or national inequality). Understanding such differences, and developing generalized theory to explain such differences, is important. Research in low and middle income countries is especially acutely needed.

MAKING SOCIAL EPIDEMIOLOGY RESEARCH MORE POLICY-RELEVANT

Many policies that are not explicitly related to medical services also influence health, as discussed in more detail in Chapter 15. Recent years have seen a marked increase in attention to these policies. Some, though not all, of this work has occurred within epidemiology or via inter-disciplinary collaborations between epidemiologists and other disciplines. Calls for formal Health Impact Assessments of policies, projects, and programs across all sectors—including food, environment, transportation, and housing—reflect this recognition (72, 73). Nonetheless, much of the research within the catchment area of "social epidemiology" seems better positioned to collect dust in under-read academic journals than to guide the thinking of influential decision-makers.

The success of "translation" depends on attending to this process in the early phases of research. As discussed above, continued focus on causal inference and improved research methods to support inferences are certainly a high priority. Good research design is the foundation of policy relevance. Other considerations may also help increase the impact of social epidemiology. In the focus on causal inference, it is sometimes overlooked that prediction can be valuable *even when causality is unclear.* Prediction models, used to create patient "risk calculators" and guide clinical decision making, have a long history in epidemiology, but are generally less emphasized in social epidemiology. Rather, we have focused on showing that social determinants were causal. Figure 12.3 (reproduced from (44)) lays out a model for Social Epidemiology Translation (SET). The SET model focuses on strategies to use evidence from social epidemiology to improve quality of clinical care. It highlights the potential value of prediction in social epidemiology: Identifying vulnerable people and vulnerable periods in people's lives can help improve the services those individuals receive, even when the markers used to identify these individuals are not themselves causal. Evidence indicates residents of low-income communities have higher post-stroke mortality rates, and are less likely than individuals in high-income communities to receive post-stroke rehabilitation services (74). This result may justify longer hospital stays, more frequent post-discharge follow-ups, and a greater allocation of case-management services for stroke survivors who live in low-income neighborhoods. Knowing whether the low-income community is causal or merely a predictive marker is not necessary to guide action; it is possible to reduce many inequalities by using strategies that do not entail changing the social factor. Fundamental cause theory would suggest that such approaches focus on proximal mediators and that they will never eliminate inequalities in health (75). This may well be true, but this observation does not diminish the importance of effective, life-saving strategies focused on proximal factors. Using social determinants evidence to help more effectively interrupt the effects of proximal factors on health may reduce the magnitude of inequalities and improve the health of at least some disadvantaged populations.

When we *do* focus on causal inference, a major challenge in social epidemiology has been to operationalize exposures with measures that closely correspond to factors that can be directly manipulated by policy changes. We often find a major discrepancy between the social exposure as included in our regression models, and the social exposure as modified by a feasible policy. For example, "income" is used to predict health, without defining the source of income. Current evidence suggests that income from wages, lotteries, or annual benefit payouts likely have different consequences for health (more evidence on this is reviewed in Chapter 2). Even worse, as discussed above, the same income source may have different consequences for different health outcomes (24, 26). Similarly, neighborhood environment is often measured with an index of socioeconomic disadvantage, but such indices have almost no correspondence with plausible social interventions or actual housing policies. Many of the most important policies are not set by legislation, but arise from executive branch decisions on how to implement a policy. For policy relevance, we must evaluate the effects of plausible policy decisions. However, identifying the range of plausible policies requires a fair amount of familiarity with the institutional details of how a resource (e.g., public housing, or transportation) is regulated. Many social policies with huge impacts on where people live are the topic of ongoing policy debates, but receive much less research attention in social epidemiology. Examples include regulations regarding availability of, and restrictions on, government vouchers providing rental subsidies to impoverished families; duration and magnitude of unemployment benefits; zoning and subsidy efforts to influence the introduction of new

resources such as grocery stores to impoverished neighborhoods; and regulations on home foreclosure protocols, for example, waiting period laws (76).

Epidemiologic research typically focuses on hypothesis testing (does exposure X increase outcome Y?), with much less emphasis on assessing the population impact of exposures or proposed policies to change population exposure to risk factors. Including estimates of population health impact and financial implications improves the usefulness of research on social determinants of health. A simple step in this direction is to include absolute effect estimates alongside relative effect estimates. This inclusion involves reporting the number of excess cases among the exposed compared to the unexposed, rather than merely calculating the ratio of rates in the exposed versus the unexposed (77). Despite controversies regarding attributable risk and attributable risk fractions (or etiologic fractions), such estimates are very useful for evaluating the overall public health importance of exposures. Attributable risk estimates, preferably based on realistic differences in population exposure, should be reported alongside honest descriptions of uncertainty due to statistical imprecision or causality (77). Common exposures with small relative effects may nonetheless have huge population effects. Because many exposures that fall within the purview of social determinants are very common, presenting attributable risk estimates highlights the potential importance of addressing social inequalities in health. In some cases, the long-term health consequences of the intersection of multiple determinants of health may best be estimated via complex systems models (78, 79). Such tools could help guide accurate and effective Health Impact Assessments, so that in debates about programs and policies—even those not directly related to health—decision makers will have useful information about likely health consequences of the proposals (72).

In addition to population health impact, policymakers need to understand the short- and long-term costs and benefits of proposed interventions. Many policy decisions must rely on quantitative assessments of the net costs of proposed policies. Evaluations of early childhood interventions may have been particularly influential because they were accompanied by cost-benefit estimates indicating that every dollar spent returned many more dollars in later years via improved incomes and saved government expenditures (80, 81). New initiatives using public-private partnerships to finance prevention programs will only be viable if we can provide rigorous evidence that such prevention programs save money in the long run (82). Social Impact Bonds (SIBs) are programs intended to address the apparent under-investment of government in prevention programs with short-term costs but long-term net savings. SIBs engage private investors to support social programs, with a government payout guaranteed only if the program achieves a pre-determined measurable goal, such as reduced re-incarceration rates or reduced hospitalization rates for high-risk populations. Recent comment noted that the primary barrier to developing SIBs was "identifying interventions with sufficiently high net benefits to allow investors to earn their required rates of return" (83). Similar concerns motivate government actions. In the United States, for example, executive branch agencies are *required* to conduct a cost-benefit analysis for all major new regulations, and these analyses are reviewed by the President's Office of Management and Budget (84). Similarly, the Congressional Budget Office scores proposed legislation for anticipated costs or savings to the federal budget; such evaluations are used to decide whether policies are consistent with spending caps (85). Although cost-benefit analyses are often not the deciding factor in a regulatory decision—many net costly policies are adopted and many

net money-saving policies are never enacted—cost estimates *can* be influential (84). But we must provide cost-benefit estimates, in order for this evidence to be incorporated into OMB estimates.

EPIDEMIOLOGIC IDEAS FOR BETTER HEALTH POLICY EVALUATIONS

Although examining health outcomes seems a natural extension of policy research on socioeconomic outcomes, research approaches that are standard in economics and related disciplines often encounter problems for health outcomes. These difficulties relate to measurement of health and potentially heterogeneous effects on different health outcomes; distinctions between health and health services; etiologic periods, given the behavioral and physiologic pathways plausibly linking social policies to health; and statistical power and imprecision.

MEASURING HEALTH

Measuring health is difficult. In general, although many domains of health are positively correlated, the correlations are modest. Furthermore, despite efforts to create a single summary measure of health, it is generally not obvious how different measures of health should be weighted. For example, both self-reported physical limitations and objectively measured physical capacities are important health outcomes, but they are not highly correlated (86, 87). Individuals with a great deal of psychological morbidity may have no physical disease (86, 87). Individuals with physical impairments may nonetheless live in environments in which they experience no effective disability. These challenges relate to measuring both population health and individual health. For example, extending population life expectancy may increase disease prevalence in the population. To inform health relevant policies it is important to distinguish effects of social policies on different health outcomes. Interventions beneficial for one outcome may well be harmful for another, and the relative merits of the intervention depend on the prevalence of the different conditions. For example, there is evidence that tobacco use decreases risk of Parkinson's disease (88), but tobacco use increases risk of heart disease (and many other conditions). There are approximately 22,000 Parkinson's disease deaths in the US each year, but about 600,000 deaths due to heart disease. Even if smoking completely eliminated Parkinson's, and had only a small impact on heart disease, smoking would still have a net adverse effect on population health. In fact, smoking is estimated to cause nearly half a million annual deaths in the US (89). Measurement of specific domains of health is important in health research, and for especially hard-to-assess outcomes, such as mental health or disability, years of research have been devoted to developing good measurement instruments. Ignoring these measures in favor of new outcome measures makes it harder to integrate findings with prior research, and risks using measures with less validity and reliability. For example, original MTO findings were published with a non-standard operationalization of psychological distress and reported no significant adverse effects on boys' distress levels. However, the non-standard operationalization of the mental health measure dichotomized items

and introduced misclassification, compared to the standard approach. Analyses using the standard operationalization demonstrated that the intervention in fact had statistically significant adverse effects on boys' psychological distress in this sample (57).

STATISTICAL POWER

Measuring health well is often expensive and time-consuming, and there is typically a direct tradeoff between quality of measurement and study sample size. Because of this, modern epidemiology places great emphasis on interpreting the confidence interval for an association, rather than statistical significance tests. We should anticipate that, for many important research questions, meta-analyses will be necessary to provide informative effect estimates; findings from individual studies are often so imprecisely estimated as to be consistent with either large benefits or large harm. Statistical power is emphasized much more in health research than in disciplines with huge data sets, or even census data, in which null findings are typically very precisely estimated. The Oregon Medicaid experiment described earlier illustrates this. The study assigned Medicaid eligibility by lottery, and results were therefore hailed as providing unbiased estimates of the effects of providing insurance coverage. Findings are often summarized as evidence of mental health benefits but not of physical health benefits. However, the confidence intervals for most physical health effect estimates include values that would have been of substantial public health impact (46).

ETIOLOGIC PERIODS

Individuals who are disadvantaged in childhood are commonly disadvantaged in later life. In addition to extending across the lifecourse, poverty and other forms of hardship commonly extend across generations, as well. Social epidemiologists hypothesize that physiologic manifestations of social disadvantage emerge at every age. Social inequalities have been documented for many early-life outcomes (see Chapters 2 and 14 on physiologic mechanisms for further discussion). Most theoretical frameworks suggest that part of the long-term health effects of social disadvantage in one period are mediated by a sequence of behavioral and physiologic changes that may manifest in full-blown disease only much later on. Because of the long-term physiologic embedding of disadvantage, analyses that condition on health have also implicitly conditioned mediating pathways between prior disadvantage and subsequent health. For example, any effects of early adult SES on late onset dementia are very likely partially mediated by cognitive functioning in late middle age. Controlling such analyses for cognitive assessments in middle age will likely underestimate causal effects of early adult SES on later dementia risk. Similar concerns relate to estimating effects on cardiovascular outcomes, while adjusting for mid-life cardiovascular risk factors (e.g., body mass index, hypertension). Even more subtle, but of potentially equal importance, conditioning on current health may induce selection bias; for example, impoverished adults with health equivalent to that of wealthy adults likely had some other resources that outweighed the adverse effects of poverty. Finally, because social conditions must be processed through an accumulation of physiologic changes before manifesting as a diagnosable clinical condition, changes in social policies are unlikely to induce changes in most "hard" health outcomes in the short term.

The correlation of social conditions across the lifecourse introduces tremendous difficulty identifying the relevant time period for interventions. Misspecifying the temporal link between the exposure and the health outcome can lead to very severe bias in even highly conservative analytic approaches such as fixed effects models. For example, changes in cigarette smoking do not predict any beneficial *short-term* changes in lung cancer risk, but may prevent cancer risk 30 years later.

CONCLUSION

Research on social determinants of health is increasingly using policy changes both to evaluate the effects of the policy and to evaluate causal hypotheses about social resources influenced by the policy. Both applications are important contributions to enhancing the relevance of social epidemiology and achieving the final goal of improving population health (90). Evidence on causation gained through evaluating the health effects of policies lays the groundwork for future translation. To have the greatest relevance, social epidemiology must borrow tools from other disciplines, including analytic and design approaches to strengthen causal inference, tools to evaluate population health impact, and cost-benefit analyses.

REFERENCES

1. Strully KW, Rehkopf DH, Xuan Z. Effects of Prenatal Poverty on Infant Health. Am Sociol Rev. 2010;75(4):534–62.
2. Sanbonmatsu L, Ludwig J, Katz L, Gennetian L, Duncan G, Kessler R, et al. Moving to Opportunity for Fair Housing Demonstration Program Final Impacts Evaluation 2011 4/16/2012.
3. Chandra A, Vogl TS. Rising up with shoe leather? A comment on "Fair Society, Healthy Lives" (The Marmot Review). Soc Sci Med. 2010;71(7):1227–30.
4. Berkman LF. Social Epidemiology: Social Determinants of Health in the United States: Are We Losing Ground? Annu Rev Public Health. 2009;30(1).
5. Clark D, Royer H. The Effect of Education on Adult Mortality and Health: Evidence from Britain. Am Econ Rev. 2013;103(6):2087–120.
6. Krieger N, Rehkopf DH, Chen JT, Waterman PD, Marcelli E, Kennedy M. The fall and rise of US inequities in premature mortality: 1960–2002. PLoS Med. 2008;5(2):e46.
7. Avendano M, Glymour MM, Banks J, Mackenbach JP. Health Disadvantage in US Adults Aged 50 to 74 Years: A Comparison of the Health of Rich and Poor Americans With That of Europeans. Am J Public Health. 2009;99(3).
8. Avendano M, Kunst AE, van Lenthe F, Bos V, Costa G, Valkonen T, et al. Trends in socioeconomic disparities in stroke mortality in six european countries between 1981–1985 and 1991–1995. Am J Epidemiol. 2005;161(1):52–61.
9. van Hedel K, Van Lenthe F, Mackenbach J. The Contribution of Larger Educational Inequalities in Mortality to the National Mortality Disadvantage of the United States. A Comparison with Seven Western European Countries. Population Association of America; New Orleans, Louisiana 2013.

10. Liu SY, Linkletter CD, Loucks EB, Glymour MM, Buka SL. Decreased births among black female adolescents following school desegregation. Soc Sci Med. 2012;74(7):982–8.
11. Liu SY, Chavan NR, Glymour MM. Type of High-School Credentials and Older Age ADL and IADL Limitations: Is the GED Credential Equivalent to a Diploma? Gerontologist. 2013;53(2):326–33.
12. Ben-Shalom Y, Moffitt R, Scholz JK. An assessment of the effectiveness of anti-poverty programs in the United States. the Johns Hopkins University, Department of Economics, 2011.
13. Engelhardt GV, Gruber J. Social Security and the Evolution of Elderly Poverty. NBER Working Paper Series. 2004;Working Paper No. 10466.
14. Snyder SE, Evans W. The Impact of Income on Mortality: Evidence from the Social Security Notch. NBER Working Paper Series. 2002;Working Paper No. 9197.
15. Handwerker EW. What can the Social Security Notch tell us about the impact of additional income in retirement? J Econ Soc Meas. 2011;36(1–2):71–92.
16. Arno PS, House JS, Viola D, Schechter C. Social security and mortality: The role of income support policies and population health in the United States. J Public Health Pol. 2011;32(2):234–50.
17. Herd P, Schoeni RF, House JS. Does the Supplemental Security Income Program Reduce Disability among the Elderly? Milbank Q. 2008;86(1):5–45.
18. Lubitz J, Cai L, Kramarow E, Lentzner H. Health, life expectancy, and health care spending among the elderly. New Engl J Med. 2003;349(11):1048–55.
19. Currie J, Cole N. Welfare and Child Health: The Link between AFDC Participation and Birth Weight. Am Econ Rev. 1993;83(4):971–85.
20. Schmeiser MD. Expanding wallets and waistlines: the impact of family income on the BMI of women and men eligible for the Earned Income Tax Credit. Health Econ. 2009;18(11):1277–94.
21. Kenkel D, Schmeiser M, Urban C. Is smoking inferior? Evidence from variation in the earned income tax credit. Evidence from Variation in the Earned Income Tax Credit (October 31, 2011). 2011.
22. Cowan B, Tefft N. Education, Maternal Smoking, and the Earned Income Tax Credit. The BE Journal of Economic Analysis & Policy. 2012;12(1).
23. Averett S, Wang Y. The effects of Earned Income Tax Credit Payment Expansion on Maternal Smoking. Health Econ. 2012.
24. Bruckner TA, Brown RA, Margerison-Zilko C. Positive income shocks and accidental deaths among Cherokee Indians: a natural experiment. Int J Epidemiol. 2011;40(4):1083–90.
25. Costello EJ, Compton SN, Keeler G, Angold A. Relationships between poverty and psychopathology: a natural experiment. JAMA. 2003;290(15):2023–9.
26. Costello EJ, Erkanli A, Copeland W, Angold A. Association of family income supplements in adolescence with development of psychiatric and substance use disorders in adulthood among an American Indian population. JAMA-J Am Med Assoc. 2010;303(19):1954–60.
27. Wilde ET, Rosen Z, Couch K, Muennig PA. Impact of Welfare Reform on Mortality: An Evaluation of the Connecticut Jobs First Program, A Randomized Controlled Trial. Am J Public Health. 2013;103(7):e1–e5.
28. Fernald LC, Gertler PJ, Neufeld LM. Role of cash in conditional cash transfer programmes for child health, growth, and development: an analysis of Mexico's *Oportunidades*. Lancet. 2008;371(9615):828–37.
29. Gertler P. Do conditional cash transfers improve child health? Evidence from PROGRESA's control randomized experiment. Am Econ Rev. 2004;94(2):336–41.
30. Behrman J, Parker S. Is Health of the Aging Improved by Conditional Cash Transfer Programs? Evidence From Mexico. Demography. 2013;50(4):1363–86.
31. Riccio J, Dechausay N, Miller C, Nunez S, Verma N, Yang E. Conditional cash transfers in New York City: The continuing story of the Opportunity NYC-Family Rewards Demonstration. MDRC; 2013.

32. Case A. *Does Money Protect Health Status? Evidence from South African Pensions.* Boston MA: National Bureau of Economic Research, 2001 Contract No.: 8495.
33. Duflo E. Grandmothers and Granddaughters: Old-Age Pensions and Intrahousehold Allocation in South Africa. World Bank Econ Rev. 2003;17(1):1–25.
34. Cylus J, Glymour MM, Avendano M. Do generous unemployment benefit programs reduce suicides? A state fixed-effect analysis covering 1968–2008. American journal of epidemiology. 2014;forthcoming.
35. Currie J, Ray SH, Neidell M. Quasi-experimental studies suggest that lowering air pollution levels benefits infants' and children's health. Health Aff (Millwood). 2011;30(12):2391–9.
36. Schwartz J. Societal benefits of reducing lead exposure. Environ Res. 1994;66(1):105–24.
37. Needleman H. Lead poisoning. Annu Rev Med. 2004;55:209–22.
38. Rosner D, Markowitz G. Special Report On Lead Poisoning In Children-Standing Up to the Lead Industry: An Interview with Herbert Needleman. Public Health Rep. 2005;120(3):330–7.
39. Currie J. U.S. Food and Nutrition Programs. In: Moffitt RA, editor. Means-Tested Transfer Programs in the United States: University of Chicago Press; 2007. pp. 199–290.
40. Schmeiser MD. The impact of long-term participation in the supplemental nutrition assistance program on child obesity. Health Econ. 2012;21(4):386–404.
41. Meyerhoefer CD, Yang M. The relationship between food assistance and health: a review of the literature and empirical strategies for identifying program effects. Appl Econ Perspect Policy. 2011;33(3):304–44.
42. Shah SM, Carey IM, Harris T, DeWilde S, Victor CR, Cook DG. The Effect of Unexpected Bereavement on Mortality in Older Couples. Am J Public Health. 2013;103(6):1140–5.
43. Shah SM, Carey IM, Harris T, DeWilde S, Victor CR, Cook DG. Impact of Partner Bereavement on Quality of Cardiovascular Disease Management. Circulation. 2013;128(25):2745–53.
44. Patton KK, Glymour MM. In Anticipation of Grief Using Insights From Social Epidemiology to Improve Quality of Care. Circulation. 2013;128(25):2725–8.
45. National Center for Health Statistics. *Health, United States, 2011: With special feature on socioeconomic status and health.* Hyattsville, MD: National Center for Health Statistics 2012.
46. Baicker K, Taubman SL, Allen HL, Bernstein M, Gruber JH, Newhouse JP, et al. The Oregon Experiment—Effects of Medicaid on Clinical Outcomes. New Engl J Med. 2013;368(18):1713–22.
47. Currie J, Gruber J. Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women. J Polit Econ. 1996;104(6):36.
48. Finkelstein A, McKnight R. What Did Medicare Do (And Was It Worth It)? J Public Econ. 2008;92:1644–69.
49. Almond D, Chay K. The Long-Run and Intergenerational Impact of Poor Infant Health: Evidence from Cohorts Born During the Civil Rights Era. Columbia University Working Paper. 2006.
50. Galizzi MM. Label, nudge or tax? A review of health policies for risky behaviours. J Public Health. 2012;1:e5.
51. Tan CE, Glantz SA. Association Between Smoke-Free Legislation and Hospitalizations for Cardiac, Cerebrovascular, and Respiratory DiseasesClinical Perspective A Meta-Analysis. Circulation. 2012;126(18):2177–83.
52. Barr CD, Diez DM, Wang Y, Dominici F, Samet JM. Comprehensive smoking bans and acute myocardial infarction among medicare enrollees in 387 US counties: 1999–2008. Am J Epidemiol. 2012;176(7):642–8.
53. Harper S, Lynch J. Trends in socioeconomic inequalities in adult health behaviors among US states, 1990–2004. Public Health Rep. 2007;122(2):177.
54. Nelson DE, Bolen J, Kresnow M-j. Trends in safety belt use by demographics and by type of state safety belt law, 1987 through 1993. Am J Public Health. 1998;88(2):245–9.
55. Howden-Chapman P, Matheson A, Crane J, Viggers H, Cunningham M, Blakely T, et al. Effect of insulating existing houses on health inequality: cluster randomised study in the community. BMJ. 2007;334(7591):460.

56. Jackson G, Thornley S, Woolston J, Papa D, Bernacchi A, Moore T. Reduced acute hospitalisation with the healthy housing programme. J Epidemiol Commun H. 2011;65(7):588–93.

57. Osypuk T, Tchetgen Tchetgen E, Acevedo-Garcia D, Earls F, Lincoln A, Schmidt N, et al. Differential Mental Health Effects of Neighborhood Relocation among Youth in Vulnerable Families: Results from a Randomized Trial. Arch Gen Psychiat. 2012;69(12):1284–94.

58. Pearl J. *Causality models, reasoning, and inference.* Cambridge [England]; New York: Cambridge University Press; 2009.

59. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. Epidemiology. 2006;17(3):268–75.

60. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Instrumental variable analysis of secondary pharmacoepidemiologic data. Epidemiology. 2006;17(4):373–4.

61. Aizer A, Doyle Jr JJ. Juvenile Incarceration and Adult Outcomes: Evidence from Randomly-Assigned Judges. NBER Working Paper. 2011.

62. Maestas N, Mullen KJ, Strand A. Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt. Am Econ Rev. 2013;103(5):1797–829.

63. Angrist J, Pischke J. Mostly harmless econometrics: an empiricist's companion: Princeton Univ Pr; 2009.

64. Lucas AM, Wilson NL. Adult Antiretroviral Therapy and Child Health: Evidence from Scale-Up in Zambia. Am Econ Rev. 2013;103(3):456–61.

65. Bollinger B, Leslie P, Sorensen A. Calorie Posting in Chain Restaurants. NBER Working Paper. 2010.

66. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology (Cambridge, Mass). 2010;21(3):383.

67. Weiss N. Can the "specificity" of an association be rehabilitated as a basis for supporting a causal hypothesis? Epidemiology (Cambridge, Mass). 2002;13(1):6.

68. Tchetgen Tchetgen E. The Control Outcome Calibration Approach for Causal Inference With Unobserved Confounding. Am J Epidemiol. 2013.

69. Bertrand M, Mullainathan S. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. National Bureau of Economic Research, 2003.

70. Ahmed AM, Hammarstedt M. Discrimination in the rental housing market: A field experiment on the Internet. J Urban Econ. 2008;64(2):362–72.

71. Thévenon O, Solaz A. Labour market effects of parental leave policies in OECD countries: Directorate for Employment, Labour and Social Affairs, OECD; 2013.

72. Collins J, Koplan JP. Health Impact Assessment. JAMA-J Am Med Assoc. 2009;302(3):315–7.

73. Joffe M, Mindell J. A framework for the evidence base to support Health Impact Assessment. J Epidemiol Commun H. 2002;56(2):132–8.

74. Kapral MK, Wang H, Mamdani M, Tu JV. Effect of socioeconomic status on treatment and mortality after stroke. Stroke. 2002;33(1):268–73.

75. Phelan JC, Link BG, Tehranifar P. Social Conditions as Fundamental Causes of Health Inequalities: Theory, Evidence, and Policy Implications. J Health Soc Behav. 2010;51(1 suppl):S28-S40.

76. Robertson C, Egelhof R, Hoke M. *Get sick, get out: the medical causes of home mortgage foreclosures.* 2009.

77. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. PLoS Med. 2007;4(10):e297.

78. Galea S, Riddle M, Kaplan GA. Causal thinking and complex system approaches in epidemiology. Int J Epidemiol. 2010;39(1):97–106.

79. Bibbins-Domingo K, Chertow GM, Coxson PG, Moran A, Lightwood JM, Pletcher MJ, et al. Projected effect of dietary salt reductions on future cardiovascular disease. New Engl J Med. 2010;362(7):590–9.

80. Belfield CR, Nores M, Barnett S, Schweinhart L. The High/Scope Perry Preschool Program Cost–Benefit Analysis Using Data from the Age-40 Followup. J Hum Resour. 2006;41(1):162–90.

81. Knudsen EI, Heckman JJ, Cameron JL, Shonkoff JP. Economic, neurobiological, and behavioral perspectives on building America's future workforce. PNAS. 2006;103(27):10155–62.

82. Center for American P, Liebman J. Social Impact Bonds: A promising new financing model to accelerate social innovation and improve government performance. Center for American Progress, 2011 2011/02//. Report No.

83. Liebman JB. Social Impact Bonds: A promising new financing model to accelerate social innovation and improve government performance: Center for American Progress; 2011. Available from: http://www.americanprogress.org/issues/open-government/report/2011/02/09/9050/social-impact-bonds/.

84. Hahn RW, Tetlock PC. Has Economic Analysis Improved Regulatory Decisions? J Econ Perspect. 2008;22(1):67–84.

85. Executive Office of the President. Circular NO. A-11, Part 7, Appendix A: Office of Management and Budget; 2013. Available from: http://www.whitehouse.gov/sites/default/files/omb/assets/a11_current_year/app_a.pdf.

86. Glass TA. Conjugating the "tenses" of function: Discordance among hypothetical, experimental, and enacted function in older adults. Gerontologist. 1998;38(1):101–12.

87. Pinquart M. Correlates of subjective health in older adults: a meta-analysis. Psychol Aging. 2001;16(3):414.

88. Wirdefeldt K, Adami H-O, Cole P, Trichopoulos D, Mandel J. Epidemiology and etiology of Parkinson's disease: a review of the evidence. Eur J Epidemiol. 2011;26(1):1–58.

89. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Borden WB, et al. Heart disease and stroke statistics—2013 update: a report from the American Heart Association. Circulation. 2013;127(1):e6–e245.

90. Hiatt RA. Invited commentary: the epicenter of translational science. Am J Epidemiol. 2010;172(5):525–7.

CHAPTER 13

APPLICATIONS OF BEHAVIORAL ECONOMICS TO IMPROVE HEALTH

Ichiro Kawachi

WHY BEHAVIOR MATTERS

In public health, the leading causes of death are often presented as "cardiovascular disease," "cancer," "accidents," and so on. But another way to frame the issue is this: The leading cause of death is behavior. Behavior matters because almost half of premature mortality (i.e., deaths before age 65) in the United States can be attributed to poor behavioral choices—including cigarette smoking, sedentarism, poor eating habits, drunk driving, risky sex, aggression, and drug abuse (1). According to Ralph Keeney (1), roughly 46% of deaths due to heart disease and 66% of cancer deaths could be avoided by helping people make better personal decisions.

The preceding paragraph tends to elicit howls of protest from social epidemiologists. People who hold strong beliefs about the social determinants of health (as I do) have a visceral reaction to the use of words such as behavioral "choice" and "personal decisions." After all, much of the rest of this book is devoted to the "upstream" drivers of health and illness. Social epidemiology emphasizes how health behaviors are shaped and constrained by the social context (see Chapter 10). To use the example of healthy eating, social epidemiologists recognize at least four different types of constraints on personal choices:

- *Information asymmetry*—food manufacturers know a lot more about what goes into making their packaged foods than consumers do, and they would prefer not to disclose some of those ingredients (e.g., trans-fatty acids) unless they are forced to do so by government regulation. Mandating the provision of nutritional information is an example of an upstream intervention to improve nutritional habits in the population.

- *Budget constraints*—people face constraints on both their time and money budgets. Time is scarce in households in which everyone of employable age is working full-time to make ends meet. When time is scarce, it may not be practical to prepare healthy, slow-cooked meals at home, and consequently the most convenient option is to eat out, order take-out, or eat in the car. The problem could be solved by hiring a cook, but few households can afford to do that.
- *Environmental constraints*—sometimes, even if one wants to eat healthily, there may not be a store or supermarket in the neighborhood that sells fresh produce. Qualitative studies conducted among residents of so-called food deserts express just as strong a preference for eating healthily compared with residents of middle-class neighborhoods (2). The problem is that resource-deprived neighborhoods are not blessed to the same extent as middle-class neighborhoods in terms of food choices in retail stores (3).
- *Social reinforcement*—put simply, we do not eat alone (or most of us don't). Our eating habits are influenced by other people to whom we are socially connected (friends, family members, coworkers), as well as by social norms (see Chapter 8). If the social norm in your workplace is to go out every night to the local pub to drink beer and eat fried snacks, then you are going to end up consuming a lot of calories and gaining weight. In other words, by joining your coworkers after work to blow off steam, you are conforming to the norm and expressing your solidarity with your group. If you decide to try to lose weight, now you are going to have to change not just your own habits, but also the behavior of your coworkers—that is, unless you don't mind being viewed as antisocial.

To the foregoing list of constraints on individual choice, we may add another:

- *The bandwidth tax*—in their book *Scarcity: Why Having Too Little Means So Much* (4) the behavioral economists Sendhil Mullainathan and Eldar Shafir argue that the scarcity of resources—that is, situations in which people find themselves short of money or time—imposes a "tax" on our cognitive functions (or the brain's "bandwidth," to use a computing analogy). When people are in the grip of scarcity, they develop tunnel vision as their attention becomes focused on the problems that are immediately before them. The bandwidth tax falls heavily on our ability to plan for the future, that is, our executive functions, including self-regulation. In short, the scarcity mindset affects those parts of our brain that are involved in helping us to delay gratification, resist temptations, plan for the future, and plan for our long-term health.

By now, I hope that it is clear that when we talk about behavioral "choices" and "decisions" in this chapter, we are not implying that individuals willfully (or even consciously) choose to engage in unhealthy behaviors. Emphatically, we are not advocating a return to the kind of victim-blaming discourse about "lifestyles" that was fashionable for a time in the 1970s, exemplified by John Knowles's infamous essay on the responsibility of the individual (5).[1] Rather, the focus of this chapter is to

1 "The cost of sloth, gluttony, alcoholic intemperance, reckless driving, sexual frenzy, and smoking is now a national, and not an individual, responsibility. This is justified as individual freedom—but one man's freedom is another man's shackle in taxes and insurance premiums. I believe the idea of a 'right' to health should be replaced by the idea of an individual moral obligation to preserve one's own health—a public duty if you will. The individual then has the 'right' to expect help with information, accessible services of good quality, and minimal financial barriers" (5: p. 59).

summarize new insights about people's judgments and choices that have emerged from the fields of behavioral economics, psychology, and neuroscience in the past three decades. As I shall argue, these insights provide a new set of tools for thinking about behavior—and how we could potentially leverage those insights to boost the success of behavioral interventions (6). For example:

- How can we more effectively counteract the information asymmetry produced by the deceptive marketing practices of the tobacco and food industry?
- How can we intervene in people's environments so that their default options point in the direction of healthier choices—or, as Richard Thaler and Cass Sunstein (7) put it, how can we be better "choice architects" to guide people's behavior?
- How can we leverage predictable decision errors to incentivize choices so that people behave in self-interested (as opposed to self-destructive) directions?

Incorporating such insights into interventions may even assist in narrowing the socioeconomic gap in health behaviors and health outcomes, for behaviors not only account for a large fraction of preventable deaths, but they also drive a significant portion of the socioeconomic gradient in health and illness. Until recently this was not the conventional wisdom taught in social epidemiology. A classic paper from the Whitehall Study of British civil servants in 1981 concluded that health-related behaviors—such as smoking, exercise, and overweight/obesity—could only account for about 40 percent of the fourfold disparity in mortality from cardiovascular disease comparing the top with the bottom of civil service grades (8). This conclusion led social epidemiologists to conjecture that the remainder of the unexplained variance was due to the psychosocial effects of social status (9). However, an updated analysis of the Whitehall II Study has led to a revision of the earlier conclusion (10). Earlier studies may have underestimated the contribution of health behaviors to social inequalities in mortality because the behaviors were assessed only at baseline and not updated during follow-up (resulting in misclassification of exposure). In the Whitehall II Study, the lowest civil service grades had 1.60 times higher risk of death from all causes than those in the highest grade. This gradient was attenuated by 42% (95% confidence interval [CI], 21%–94%) when health behaviors (assessed at baseline only) were entered into the model; but the degree of attenuation of the gradient rose to 72% (95% CI, 42%–154%) when health behaviors were entered as time-dependent covariates (10). In short, behavior matters for socioeconomic status (SES) gradients, too.

LEN'S CHALLENGE

S. Leonard ("Len") Syme, who is surely one of the most distinguished social epidemiologists of our time, has a habit of throwing out a challenge to his colleagues, namely, "Why do behavioral interventions fail?" We call this "Len's challenge," but it might also be considered one of the Grand Challenges of public health. According to Syme (11),

> the problem is that even when people know about their risk, they find it difficult to change their behavior. There are many examples that describe the failure of wonderfully

designed and executed interventions to help people lower their risk. In fact, I participated in one of them: the Multiple Risk Factor Intervention Trial. This $200 million study involved men in the top 10 percent risk category for developing heart disease. We screened 500,000 men in 22 cities and selected 12,000 highly informed and motivated participants for a 6-year trial. We asked them to change their diet, take high-blood-pressure medication, stop smoking, and report frequently to the clinic. Together, we cooked low-fat meals and read labels at supermarkets. We conducted a superb intervention program, but the trial failed. After 6 years, there was no statistically significant difference in heart disease rates between our group and the control group. Few men in our group changed their behavior.

Len's challenge is more than a personal anecdote. A systematic review by Pennant et al. (12) sought to assess the effectiveness of community-based programs for prevention of cardiovascular disease (CVD). These programs were multifaceted interventions employing combinations of media, screening, counseling activities, and environmental changes to address cardiovascular risk behaviors (such as smoking, hypertension, physical activity, and diet). The authors conducted a comprehensive search of databases and relevant websites from January 1970 to mid-July 2008. Only the most rigorously designed programs were admitted, that is, studies that involved both a control group, as well as before/after design to examine changes in cardiovascular disease risk. Net changes in CVD risk factors were used to generate an overall index for net change in 10-year CVD risk. After extensive searches, the authors identified 36 relevant community programs that met their inclusion criteria. What did the authors find? The average net reduction in 10-year CVD risk across studies was 0.65%—a statistically significant but modest gain (to put it charitably). A subset of seven studies was able to incorporate "hard" outcomes including changes in CVD/total mortality rates; in none of these studies did the investigators report statistically significant reductions.

When we turn to population trends in health behaviors, we are either treading water or in some instances, losing ground. King and colleagues (13) examined national trends in five health behaviors among adults aged 40–74 years, based on analyses of the National Health and Nutrition Examination Survey 1988–1994 and the National Health and Nutrition Examination Survey 2001–2006. For three of the five behaviors, adherence to a healthy pattern has decreased among the general American public during the last 18 years—namely, intake of five or more servings of fresh fruits and vegetables/day; regular exercise >12 times/month; and maintaining healthy weight (body mass index [BMI] 18.5–29.9 kg/m2) (13). During the past two decades, the prevalence of adult obesity (BMI ≥ 30 kg/m2) increased from 28% to 36% ($p < 0.05$); physical activity 12 times a month or more decreased from 53% to 43% ($p < 0.05$); smoking rates remained unchanged (26.9% to 26.1%); eating five or more fruits and vegetables a day decreased from 42% to 26% ($p < 0.05$), and moderate alcohol use increased from 40% to 51% ($p < 0.05$). Adherence to all five healthy habits dropped over the same period from 15% to 8% ($p < 0.05$). Furthermore, although adherence to a healthy lifestyle was lower among minorities, adherence decreased even more among non-Hispanic whites over the period. Individuals with a history of hypertension, diabetes, or cardiovascular disease were no more likely to be adherent to a healthy lifestyle than people without these conditions (13).

WHY IS BEHAVIOR CHANGE SO DIFFICULT?

By now, we have a fairly robust evidence base on which to recommend what behaviors people ought to follow in order to lead a long and healthy life—namely, don't smoke, eat less, stay lean, exercise regularly, don't drink and drive, floss after meals, sleep at least seven hours, and so forth. The problem is we have not yet solved the problem of how to help people follow public health advice. Why is behavior change so difficult? There are at least three answers to that puzzle.

The first reason is that many behavior change interventions are overly focused on intraindividual factors (such as improving knowledge and intentions) and fail to give sufficient attention to interpersonal influences such as local norms and network influences (see Chapters 7 and 8) or extraindividual factors such as environmental barriers (see Chapter 10). Social epidemiology has endeavored to raise awareness of the *embeddedness* of health behavior within a social context (see Chapter 10). Sustained behavior change is unlikely to occur until and unless we recognize the social context that drives individual motivation to change. For example, Sorensen and colleagues (14) have persuasively argued that no matter how much effort we expend on distributing informational leaflets and other health education strategies, we are unlikely to make a dent in the high prevalence of smoking in blue-collar manufacturing worksites. The reason is because many workers are also exposed to other chemical and physical hazards in their workplace, that is, many workers face the double jeopardy of exposures to occupational carcinogens and personal risks such as smoking. In other words, they may be behaving quite "rationally" in concluding: "What's the use of giving up smoking if I turn around and get exposed in my job to the same carcinogens that are present in cigarette smoke?" Based on this insight, Sorensen et al. (14) developed a novel intervention model that integrated health promotion and health protection through (1) joint worker-management participation in program planning and implementation, (2) consultation on worksite changes, and (3) educational programs targeting health behavior change. In a rigorously designed cluster-randomized trial involving 24 worksites, they showed that the joint occupational safety/health education intervention was superior to the traditional educational strategy in persuading blue-collar workers to quit. In fact, they almost eliminated the occupational class disparity in smoking as a result of the novel intervention.

Sorensen's WellWorks intervention illustrates what economists have called the principle of *complementarity* (15), that is, making investments to improve health in one domain can increase the marginal benefit of investing in an unrelated area. For example, when widespread childhood immunization campaigns were carried out in sub-Saharan Africa, local health workers noted a subsequent rise in breast-feeding rates (15). In short, investing in the future survival of infants increased maternal motivation to breast-feed their babies. In an analogous manner, investing in the health and safety of the workplace can motivate workers to begin investing in their own health. The two investments are complementary.

A second reason why it is challenging to change people's health behavior is that almost every piece of advice that we dispense in public health turns out to have a counter party, and "they" stand to make a profit by persuading people to do the opposite. Thus, when we consider the range of behaviors that we advocate—not smoking, eating less, drinking in moderation, stop watching television, and going to the gym—almost every one of these behaviors are connected to industries

which make money by convincing people to do the opposite. Moreover, these groups have advertising budgets that dwarf public health budgets. However, as we shall see in a moment, it is not just a matter of the differential in the size of marketing budgets; they also advertise using different techniques than we do.

Last but not least, the third reason why behavior change is difficult is that as a description of individual decision-making, standard theories of behavior (on which behavior interventions are modeled) need some tweaking. Take two workhorse theories of behavior standardly taught in behavioral science programs—the theory of reasoned action (TRA) and its extension the theory of planned behavior (TPB) (16). As implied by their titles, these theories are predicated on the notion that individuals reason and plan their way to behavior. They are examples of a broader class of *expectancy-value theories*, in which behavior is held to be a function of the expectancies one has about the behavior (i.e., subjective probabilities of alternative outcomes) and the value (or utility/disutility) that one has assigned to the outcomes of the behavior. The approach predicts that, when the individual is facing a decision (e.g., "Do I quit smoking or continue smoking?"), the behavior chosen will be the one with the largest combination of expected success and value. Accordingly, one target in designing a behavior change intervention based on TPB is to attempt to modify the smoker's beliefs about the benefits and costs of her behavior.

One of the core claims of TRA and TPB (and its latest iteration, the integrated model) is that behaviors are preceded by *intentions* to perform that behavior. Intentions are in turn shaped by one's attitudes toward the behavior (the cost/benefit of the behavior and outcome expectations), perceived norms about the behavior (whether others approve/disapprove of the behavior, weighted by motivation to comply), and control beliefs (self-efficacy or perceived behavioral control). A striking claim made by the developers of the theory is that "there is only a limited number of variables that need to be considered in order to predict, change, or reinforce a given behavior" (17). How well does the theory perform in explaining health behavior?

In fact expectancy-value theories are quite good at explaining people's intentions to perform a behavior; however, it is quite another thing whether our intentions predict actual behavior. The majority of empirical demonstrations of the intention-behavior relation involve observational studies that make causal inference difficult. In order to examine whether *changes* in behavioral intention engender *changes* in behavior, Webb and Sheeran (18) meta-analyzed 47 experimental tests of intention-behavior relations. The meta-analysis showed that a medium-to-large change in intention (the effect size measured by Cohen's d statistic, defined as the difference between two means divided by a standard deviation = 0.66) leads to only a modest change in behavior (d = 0.36).

There are two important reasons why intentions do not reliably predict behavior. First, many (possibly most) behaviors skip or bypass conscious intentions; they are instead automatic or habitual, or swayed by momentary influences such as emotions. Second, even when we do form an intention to perform a behavior (such as going to the gym every day), we often fail to follow up on that intention ("Not today—I'll go to the gym tomorrow"). Our intentions are often unreliable; moreover, they are systematically and predictably unreliable as the temporal interval increases between when we form an intention to the moment we confront the actual behavioral choice. In the jargon of behavioral economics, our preferences are not stable or dynamically consistent. We discuss each of these problems.

HEURISTICS AND BIASES

Contrary to the image of people as rational calculating machines, people behave most of the time in automatic ways, basing their judgments and decisions on mental shortcuts (called *heuristics*). It makes evolutionary sense that we do not go about our day pausing to make thousands of instantaneous cost-benefit calculations before choosing to engage in a particular behavior. Frankly that would be very inefficient and a waste of mental energy. Accordingly, much of our judgment and decision-making is fast and automatic, and this frees up our time to engage in more important pursuits. However, these mental rules of thumb can occasionally lead us disastrously astray. In a 1974 *Science* paper, Amos Tverksy and Daniel Kahneman (19) described several examples of heuristic biases in their "shot across the bow" challenging of the idealized notion of human beings as hyperrational calculating machines. Since Tversky and Kahneman's seminal paper, dozens of other kinds of heuristics and biases have been added to the collection (it's become a veritable museum of curiosities), but the ones we shall dwell on here as being the most relevant to health decision-making are the affect heuristic and anchoring bias.

THE AFFECT HEURISTIC

The affect heuristic refers to our tendency to base a judgment or decision on the basis of an emotional reaction rather than a cold calculation of risks and benefits. The journalist Dan Gardner (20) calls it the good/bad rule of thumb, that is, when we have a favorable emotional reaction to something we have a tendency to judge it as "good" and vice versa. Importantly, from the point of view of explaining health behavior, the heuristic leads us to judge the riskiness of a choice as being low when we perceive its benefits as being high. Put another way, there is an inverse relationship between perceived risk and perceived benefit (21). In the real world this does not make sense. In the real world, there tends to be a positive relationship between risk and benefit; for example, smokers persist in their habit at considerable risk to themselves because it is enjoyable. If smoking were both high risk and low benefit, we would not have any problem in persuading people to stop, or society would encounter little resistance in outlawing it. Experiments conducted by Finucane et al. (21) have demonstrated that the good/bad rule can be manipulated. Thus, when information was provided to favorably alter an individual's affective evaluation of an item (their example was nuclear power), it resulted in a systematic change in people's risk/benefit judgments. Importantly, when people were manipulated so that their liking nuclear power was increased, their assessment of the risks of nuclear power was lowered, *even though no additional information was provided to them concerning the risks*. This is precisely the reason why the marketing of tobacco products appeals to positive emotions such as pleasure, joy, and happiness. Frequently, the advertisements have nothing to do with smoking. The images are of people tobogganing down a powdery slope or dancing outside a poolside cabana; often the individuals are not even depicted in the act of smoking. Yet if the affect heuristic works as intended, the advertisements are designed to generate a positive emotion about the product and thereby lower the consumer's assessment of risks.

In public health, the use of the fear appeal is an attempt to leverage the good/bad rule in the opposite direction, that is, if we can generate a negative emotion in the consumer (such as fear or

anxiety), we might succeed in raising risk perceptions. Unfortunately, a meta-analysis of the use of fear appeals in public health messaging campaigns has suggested that they have somewhat limited efficacy in motivating behavior change (22). One reason is because fear appeals can backfire if target audiences do not believe they are able to effectively avert a threat. Exposing consumers to frightful images of blackened lungs can result in defensive reactions such as denial or avoidance. In short, there is an unavoidable asymmetry in the marketing strategy of the tobacco industry versus that of public health. "Their" side gets to appeal to positive, "feel good" emotions, whereas "our" side has to resort to scary, negative emotions. But is this strictly true? Is fear the only emotion that is available to the public health campaigner?

In fact, the theory of the affect heuristic suggests that there may be a much broader palette of emotions that can potentially be manipulated to induce changes in risk perceptions. There is a tendency in public health to immediately reach for fear as the principal weapon of persuasion. Theories of behavior change such as the health belief model (23) suggest that fear ought to raise the levels of perceived severity and susceptibility. However, more recent insights from behavioral theory suggest that we do not always need to scare the bejesus out of the public in order to achieve a result. According to Lerner and Keltner's (24) appraisal tendency theory, the influence of emotions on consumer judgment and choice depend not just on the valence of the emotion (negative vs. positive), but the *specific* type of emotion can also induce different risk assessment and action tendencies. For example, fear and anger (along with anxiety, sadness, and disgust) are classified as negative emotions; yet experimentally, they have been shown to be associated with opposite appraisals about controllability. In brief, fear appeals tend to lead to an appraisal of negative events (such as getting cancer) as being unpredictable and beyond the control of the individual, whereas anger tends to lead to an appraisal of negative events as being more predictable and under human control (24). This kind of nuanced insight helps to explain why appeals to anger (against industry manipulation) have proved quite effective in some antitobacco campaigns, such as the "Truth" campaign led by the American Legacy Foundation (25).

More nuanced understanding of specific emotions also holds promise for a more systematic approach to designing health communication messages. For instance, in accordance with the Family Smoking Prevention and Tobacco Control Act (2009), the Food and Drug Administration proposed new graphic warnings that will take up 50% of surface area of the front of cigarette packs. A noteworthy feature of these proposed warnings is that they did not just target fear; the rotating warnings included the full palette of emotions, from positive (e.g., pride in having successfully quit) to negative (e.g., sadness about causing harm to loved ones via secondhand smoke). Infamously, the FDA proposal was blocked by a lawsuit by US tobacco companies. In siding with the tobacco manufacturers, the presiding US District Judge ruled that "it is abundantly clear from viewing these images that the emotional response they were crafted to induce is calculated to provoke the viewer to quit, or never to start smoking—an objective wholly apart from disseminating purely factual and uncontroversial information."[2] In other words, the FDA graphic warnings were likely to succeed by fighting fire with fire. Conversely, so long as government warnings are restricted to emotionally neutral and "factual" messages about the risks of smoking, it is like stepping into a boxing ring with one hand tied behind one's back.

2 "Court Says FDA Warnings Violate First Amendment," Constitutional Law ProfBlog, March 1, 2012. Accessed at: http://lawprofessors.typepad.com/conlaw/2012/03/court-says-fda-warnings-violate-first-amendment.html.

DUAL PROCESSING THEORY

Heuristics are a feature of what Kahneman (26) calls System 1 thinking. In behavioral economics and neuroscience, human judgments and choices reflect the dual influence of the two systems: System 1 (intuition) and System 2 (reasoning). Thought processes in System 1 tend to be intuitive, fast, automatic, and often emotionally charged. By contrast, thought processes in System 2 tend to be reflective, slow, effortful, and deliberate. The distinction is important because in the example cited above, tobacco warning labels were legally constrained to target System 2 processes (the presentation of "factual" information to the consumer), whereas most of the persuasive tactics used by the tobacco industry target System 1. Thus the two sides end up targeting different parts of the human brain. When a tobacco manufacturer takes out a full-page color advertisement in a magazine depicting a group of people "alive with pleasure," they are targeting System 1 (via the affect heuristic). By contrast, existing government warnings in the United States—which appear inconspicuously at the bottom of such ads—ask the smoker to imagine the consequences of their behavior in the distant future, such as developing emphysema; in other words, they appeal to System 2.

Some models of individual health behavior also tend to give short shrift to the influence of emotions in decision-making. For example, the textbook illustration of the integrated behavior model incorporates emotion on the far left-hand side of the causal pathway, buried in a long list of "Background Influences" on behavior (which also includes "demographic factors and culture" and "past behavior") (16). Emotions are thereby relegated in the integrated model to the side-show in terms of what "really" determines behavior. In other accounts of choice in the decision sciences, the story is the same—emotions are at best treated as by-products of the decision-making process, that is, as epiphenomenonal (27). As remarked by Loewenstein et al. (27), "virtually all current theories of choice under risk or uncertainty are cognitive and consequentialist. They assume that people assess the desirability and likelihood of possible outcomes of choice alternatives and integrate this information through some type of expectation-based calculus to arrive at a decision" (27). On occasion, emotions find their way into expectancy-value theories as an input into the decision-making process. For example, the likely emotional consequences of a bad choice (regret) can be incorporated into the expectations calculus at the point of decision. However, as Loewenstein et al. (27) argue, there is a crucial distinction between such *anticipated* emotions and *anticipatory* emotions. Anticipated emotions are just another piece of cognition (System 2 thinking), whereas anticipatory emotions are those felt *at the time of decision-making*, which can influence cognitive evaluations and choices. A key insight of behavioral economics is that these ambulatory emotions ought to be taken into consideration as both a direct input into risk/benefit judgments (the "risks as feelings" hypothesis) as well as a direct influence on consumer choice. Contrary to the predictions of expectancy-value theories, our incidental emotions frequently cause us to skip cognitive processes altogether in deciding to perform a given behavior. As a consequence, these behavioral models are "somewhat crippled emotionally, and thus detached from the emotional and visceral richness of life" (28).

ANCHORING BIAS

Another type of heuristic that is highly relevant to health behaviors is the anchoring and adjustment heuristic (19). In a famous demonstration of this effect, Tversky and Kahneman asked a

group of students in class to guestimate the percentage of African nations belonging to the United Nations. Before the students wrote down the answer, the professors whirled a giant roulette wheel positioned at the front of the classroom. After observing the number on which the pin settled, the students were asked to indicate whether the true answer to the quiz was higher or lower than that number, and then to write down their best guestimate. The trick in this experiment was that unbeknownst to the students, the roulette wheel was rigged to land only on two numbers—either 10 or 65. In the classroom in which the roulette landed on the number 10, the students' median estimate of the percentage of African nations in the UN was 25%; in the class in which the roulette landed on the number 65, the median estimate was 45% (19). The same kind of bias has been replicated in auction experiments in which simply asking people to write down the last two digits of their social security number at the top of a sheet of paper (on which they also wrote down their bids for a box of chocolate, a bottle of wine, and so on) influenced the amount of money they were willing to pay for the various auction items. For instance, people whose social security numbers ended in the range 80–99 were likely to bid twice as high for a box of chocolate compared to people whose numbers ended in the range 00–10 (29).[3]

The anchoring heuristic thus refers to the phenomenon whereby judgment and choice become tethered to irrelevant information presented at the same time. An extension of this phenomenon in the public health realm occurs when our eating habits become anchored to external cues, such as the serving size of food. Brian Wansink (30) has pioneered the notion of "mindless eating." It is a common myth that people regulate their food intake based on how hungry they are feeling and how good the food tastes. According to Wansink's experiments, people tend to eat whatever is given to them; we eat more when the food is served in larger containers—even if it tastes awful. An amusing demonstration of this phenomenon was made by Wansink's "popcorn experiment" (31). In this experiment, moviegoers were provided with either a medium-sized bucket of popcorn or a large-sized bucket. The catch was that regardless of the serving container, all the popcorn was five days old and (in the words of the researchers) "tasted like Styrofoam packing."[4] After the subjects watched the movie, the researchers collected the popcorn containers to measure how much had been consumed. Conforming to the anchoring hypothesis, customers who were given the large-sized bucket consumed on average 55% more popcorn than those given the medium-sized bucket (equivalent to ≈ 170 extra calories, or 21 more dips). Other experiments by Wansink (30) corroborate that anchoring is a robust phenomenon in food intake; for instance, the larger the bottle of cooking oil that we purchase, the greater the quantity of oil we pour to fry chicken, or the larger the packet of pasta we buy, the more strands of pasta we use to prepare a meal. It is the reverse of this same logic that prompted Mayor Bloomberg to attempt to restrict the sale of super-sized sugar-sweetened beverages in New York City in 2012.

The concept of portion control leverages the anchoring heuristic so that consumer behavior is nudged in the direction of healthy eating. In a randomized controlled trial of 130 obese patients with type 2 diabetes, Pedersen et al. (32) assigned the subjects to either daily use of a commercially available portion control plate for 6 months or usual care (dietary advice). The dinner

3 Afterward, when the bidders were asked whether writing down their social security numbers might have influenced their bid price, 100% answered "No way!"

4 A good manipulation check was provided by the fact that afterward some outraged subjects asked for their money back, forgetting that the popcorn was provided free.

plate in the intervention group was calibrated to a size so that the food heaped on top of it would amount to about 880 kcal. (for men) or 650 kcal. (for women). The surface of the plate was further partitioned into different sections by a tape, so that there were separate sections for piling on the vegetables as opposed to the meats. Similarly, the breakfast bowls provided to the intervention group were scored with different "water lines" to indicate the recommended intakes of cereal and skim milk. After six months of the trial, patients in the intervention group lost significantly more weight than control subjects (mean+/-SD, 1.8%+/-3.9% vs. 0.1%+/-3.0%, $p = 0.006$). Compared with controls, patients in the intervention group were more likely to experience a decrease in their diabetes medications at 6 months (26.2% vs. 10.8%, $p = 0.04$).

DEFAULT OPTIONS AND NUDGES

Our behavioral habits are anchored by features of our environment. If it is the societal norm that the size of the standard dinner plate is 11.5 inches in diameter, then our meal portions will be anchored to that size. By one estimate the surface area of the average dinner plate in American society has increased by 36% since 1960 (30), and therefore we should not be surprised that serving sizes have risen in lockstep. For example, in the 2006 edition of the *Joy of Cooking*, the serving size of some entrees had increased by as much as 42% compared to the same recipes in the first edition of 1931 (33). The flip-side of this finding is that we can promote health by arranging defaults in the environment so that people can be nudged to make healthful choices. Decision-making experiments have shown that people disproportionally prefer the status quo to changing their routines. This status quo bias can be turned to the advantage of promoting healthier behavior through the use of "default options" advocated by Thaler and Sunstein (7)—also referred to as "behavioral nudges."

Julie Downs and colleagues (34) demonstrated the powerful effects of a behavioral nudge through a randomized experiment conducted in a New York fast-food sandwich restaurant. Patrons of the restaurant were offered a free meal (sandwich, side, drink) in exchange for completing a survey. The survey itself was a decoy; the real purpose of the experiment was to compare the differential impact of: (1) providing calorie information on the menu versus (2) altering the convenience of ordering healthy versus unhealthy sandwich orders. The behavioral nudge itself was deceptively simple. All customers were given a menu binder displaying the options for different "submarine" sandwiches sold in the restaurant. In the "high-calorie" conditions, customers were given a menu binder in which the first page displayed the highest calories items. At the bottom of the page, in large print, the subjects were informed that "additional subs (i.e., lower-calorie subs) are available in the back of the binder." For customers randomized to the "low-calorie" condition, the menu binder was organized in the opposite way, that is, the healthy low-calorie items were featured on the front page, while unhealthy, high-calorie sandwiches were shoved to the back of the menu. The results were startling. When the default option was set to showing the healthier items first, customers ordered significantly fewer total calories (77 calories less, $p < 0.05$) compared with the unhealthy default. The effect of the nudge was bigger than printing the calorie information on the menu (in which the average customer ordered about 48.0 fewer calories, $p < 0.01$). Since requiring restaurants to post calorie information on menus is more costly than asking them

to rearrange their menus, the nudge strategy is arguably more cost-effective. In a subgroup of subjects, the informational strategy in Downs's experiment was even shown to backfire, that is, in the subset of individuals who happened to be trying to lose weight when they were enrolled in the experiment, providing them with accurate calorie information made them 76% less likely to order a low-calorie sandwich ($p < .01$). The researchers attributed this apparent perverse effect to the fact that people on a diet often motivate themselves by exaggerating the calories in the foods they eat (as a self-control device). Accordingly, providing them with accurate information may have resulted in a downward revision of calorie estimates, and hence increased intake (the "Hoagie is only 500 kcal" effect).

The potential power of default options is evidenced by another example of a simple nudge. Many people are unaware that when they order in a franchise fast-food restaurant, the cashier has been trained to automatically prompt the customer with questions like "Would you like to super-size that order?" or "Would you like to add a drink to that order?" Reversing this default, Schwartz et al. (35) conducted an experiment in a campus cafeteria by training the servers to prompt customers if they would like to "downsize" their meal portions. The experiment took place in a fast-food Chinese restaurant, where customers who ordered the high-calorie menu items (such as fried rice and *chow mein*) were asked whether they would prefer to just have one scoop instead of the usual two scoops. When consumers' self-control was activated by this simple approach, they took up the downsizing offer in 14–33 percent of cases, and they did so whether or not they were offered an additional 25-cent discount on the price of their meal. Overall, those who accepted smaller portions did not compensate by ordering more calories in their entrées, and the total calories served to them were, on average, reduced by more than 200 (35). The researchers also established that accepting the downsizing offer did not change the amount of uneaten food left at the end of the meal, so the calorie savings during purchasing translated into calorie savings during consumption. Lastly, echoing the findings of Downs et al. (34) concerning the limited effectiveness of informational strategies, the researchers found that when combined with the downsizing intervention, the posting of caloric information backfired—only 14% took up the downsizing offer compared to 21% in the absence of calorie posting. In other words, the two interventions (posting nutritional information and asking about downsizing as a default) did not have additive effects in encouraging consumers to order fewer calories.

Some people express unease with the idea that public health interventions should take advantage of default options. It sounds insidious, like consumer manipulation. But the fact of the matter is that default options are ubiquitous in our lives. It's just that most people are unaware that they are already being manipulated on a daily basis by the food industry, the retail industry, and the marketing industry. For example, when you walk into your local supermarket, most likely you enter through the door on the right-hand side of the building (as you face the front), and make your way counter-clockwise through the store. The reason is because marketers have discovered that the counter-clockwise store design leads to higher sales—something to do with right-handed people being used to looking to their right (36). As you make your way through the store, you may notice a bin full of pasta which is advertised as "Buy 3 for $3." This seems like a good deal, until you notice that in the pasta aisle, the same brand is selling for $1 a box. The "Buy 3" is just another nudge leveraging the anchoring heuristic. In the meat aisle, as you shop for the ground beef to make your Bolognese sauce, you reach for the packet labeled "85% fat-free." What does "85% fat-free" mean? It's the same as "15% fat"—the only reason it's labeled "85% fat-free" is because the beef industry has been allowed to frame it that way. Moving along,

when you pick up your 2% fat milk, don't be fooled into thinking that "2% fat" means "98% fat-free," as most people mistakenly believe. "2% fat" simply means the product is 2% fat by weight. The actual percentage of fat by calories in a glass of 2% milk is about 35%—not a lot lower than a glass of full fat milk (which is 47% fat). Finally as you check out at the counter, some supermarkets have recently stopped printing paper receipts as the default option (if you want to keep a paper record of your purchases, you have to take the extra step of asking the cashier for one). While this move is trumpeted by the retail industry as their contribution to saving the environment, it also saves them a ton of money—not just the savings from printing extra paper, but also because the majority of the time, errors at the checkout have been shown to favor the store (i.e., advertised "specials" that they neglect to input into the scanner), and because if something turns out to be not right when you get home (e.g., your eggs are broken), you are more likely to feel reluctant to return your damaged goods without a paper receipt. In a word, whether or not we decide to get on board with default options, the private sector is doing it to us all the time.

L'ENFER EST PLEIN DE BONNES VOLONTÉS ET DÉSIRS[5]

Earlier, we mentioned that there are two challenges to the claim that the most important antecedent of behavior is the intention to perform that behavior. The first challenge is that many decisions bypass conscious intentions (as when we make snap decisions based on ambient emotions, e.g., we are unconsciously swayed by the jaunty music playing in the store). The second challenge which we now turn to is that our intentions are unstable and unreliable. They are especially unreliable when there is a temporal gap between the intention and the behavior. For instance, the night before going on a business trip, Ichiro has a strong preference to stay healthy and avoid eating junk food. He resolves to eat only healthy meals while he is away from home. However, the moment he arrives at the United Airlines terminal at Logan Airport, and he sees the long queue snaking in front of the TSA screening, he experiences the draining of his willpower (i.e., a temporary weakening of his System 2 to monitor the impulses of System 1). At the same moment, he is assaulted by the smell of French fries wafting from the burger joint located at the corner of the terminal. He makes a beeline to the fast food counter. Later on, when he is sitting on the plane and recovering from the indigestion caused by the hastily scoffed French fries, he is back again to his original preference (experienced in the form of regret). This example illustrates how consumers' preferences are not dynamically stable.

One of the biggest challenges to behaving healthfully arises when the benefits and costs of a behavior fall in different time periods, also known as the problem of intertemporal choice in behavioral economics. As we pause to think about this for a moment, almost every behavior which the public health profession foists on the general public turns out to be characterized by the problem of intertemporal choice. Victor Fuchs (37) categorized these behaviors into two classes.

5 Translates as "Hell is full of good wishes and desires," by Saint Bernard of Clairvaux (c. 1150), or in plain English, "The road to hell is paved with good intentions."

First we have *investment behaviors*—such as going to the gym to get our daily dose of exercise, or flossing our teeth after every meal. In both cases, there is a cost; whether it is an economic cost or a physical/psychic cost (such as suffering on the treadmill) or an expenditure of time (such as flossing) doesn't matter; in both cases, the "pain" is now, while the benefit is reaped at some point in the future, in the form of lower dental bills or reduced risk of having a heart attack. The mirror image of these behaviors is classified as *sinful goods*, in which the fun is now and the cost comes later—for example, cigarette smoking, or risky sex, or pigging out on Krispy Kremes. In each case, the costs and benefits fall in different time periods, yielding plenty of opportunity for procrastination (in the case of investment goods) or temptation (in the case of sinful goods). If cigarette smoking led to instant death or immediate skin wrinkles, the public health profession would have little difficulty in convincing the public to cease and desist.

In economics, the problem of intertemporal choice is recognized and dealt with using the concept of *delay discounting*. Consumers are said to display positive time preference (i.e., if we had our druthers, we would prefer to enjoy something pleasurable now and put off the costs till later), and the differential ability to delay gratification for future benefit is expressed by the concept of the internal discount rate. In economic theory, discounting is modeled in terms of how large a premium a consumer will place on enjoyment nearer in time over more remote enjoyment. An individual with a high discount rate is focused more on gratification in the present, as opposed to putting it off for the future. For example, suppose that someone values eating a glazed donut now at 100 pleasure-units (let's call them *utils*). The value of delaying the consumption of the same donut till time t+1 is worth only $100 \times (1/(1 + r))^t$ utils in present terms, where r is the person's internal discount rate. If the person's discount rate for enjoying donuts is 5% ($r = 0.05$), then the net present value of putting off eating it till the next time period is equivalent to only 95 utils. Putting off the consumption for 2 time periods (t+2) makes it worth only 90 utils, and so on. According to this standard economic account of intertemporal choice, the exponential utility function declines at a constant, stable rate, implying time-consistent preferences (i.e., the individual is assumed to apply the same internal discount rate, r, to all future time periods).

However, real people are not like that. For example, suppose you were offered a choice between receiving $100 cash one year from today versus receiving $120 one year-plus-one month from today. Which would you choose? Most people would choose to wait the extra month to receive the extra twenty dollars (if you chose the $100 option, maybe you know a very good secret about where your money can earn better than 20% interest per month). How about if we change the choice—Suppose you are offered $100 cash NOW, or $120 cash one month from now. Which would you choose? If you picked $100 cash now, you are exhibiting a normal human tendency, which is that people tend to be much more impatient when the prospect of reward is immediate. This type of preference switch does not conform to the rational actor model of classical economics—after all both choice sets involve waiting an extra month to earn 20% more money. An individual who chose to wait 13 months to earn $120 ought to be indifferent to waiting an extra month to earn the same amount in the second scenario. Nevertheless, when the choice involves alternatives which are in the distant future (wait one year versus wait 13 months), we seem much more willing to be patient. The tendency for immediately available rewards to have a disproportionate impact on preferences relative to more delayed rewards is called present-focused preferences or myopia, and the discount function in this case is best described by a *hyperbolic* pattern of decay over time (as opposed to an exponential rate of decay).

THE NEUROSCIENCE OF INTERTEMPORAL CHOICE

Everything we have touched on so far—the theory of dual processing, the distinction between System 1 and System 2 processes, and the struggle between instant gratification versus delaying gratification for a larger reward—turns out to have a basis in neuroscience. In the burgeoning field of "neuroeconomics," researchers have begun to localize the brain regions associated with different choice tasks with the aid of functional MRI imaging. For example, McClure et al. (38) reported an experiment carried out among a group of subjects who were instructed to abstain from all fluids for three hours so that they were unbearably thirsty by the time they presented to the MRI laboratory. While being scanned by the fMRI machine, the subjects were offered a series of binary choices between receiving "X squirts of juice through a straw at D minutes delay" versus "X+α squirts of fluids at D+α minutes delay." For example, the subject could choose to receive 1 squirt of juice NOW or wait one minute to receive two squirts of juice, and so on.[6] The first noteworthy finding of this experiment is that the researchers found behavioral evidence in support of nonexponential discounting. In other words, the subjects were much more likely to exhibit impatience when the choice option included the prospect of receiving a squirt of juice NOW (as opposed to a choice between two delayed rewards). To wit, people tend to discount delays hyperbolically—their choices reflect big reductions in the value of the reward for the first small bits of delay, and relatively small reductions in value for subsequent increases in delay.

The second noteworthy finding of this experiment was that blood flow to the limbic area was greater for choices between an immediate reward and a delayed reward than for choices between two delayed rewards. The regions that "lit up" in response to the offer of an immediate reward included the nucleus accumbens (NAcc), subgenual cingulate cortex (SGC), posterior cingulate cortex (PCC), and the precuneus (Pcu). They seem to be the brain areas involved in System 1 processes. By contrast, areas of the brain such as the lateral prefrontal cortex and posterior parietal cortex responded similarly regardless of whether choices were between an immediate and a delayed reward or between two delayed rewards. These areas—the posterior parietal cortex (PPar), the anterior insula (Ant Ins), the posterior cingulate cortex (PCC), and the dorsolateral prefrontal cortex (Brodmann areas 9, 44, 46, and 10)—appear to reflect System 2 processes. Last but not least, the relative activation of the two sets of brain regions seemed to predict actual choice behavior. Individuals who had a preponderance of System 1 area activity over system 2 activity exhibited more impatience during the experiment. Obviously, based on the present state of knowledge, we need to be a bit cautious about pushing the brain localization too far. As Kahneman has written (26), System 1 and System 2 are "fictitious characters. . . . And there is no one part of the brain that either of the systems would call home" (p. 29). Nevertheless, we appear to be in the midst of a convergence of theories and evidence across the fields of psychology, behavioral economics, and neuroscience.

6 The design is rather analogous to Walter Mischel's famous "marshmallow experiments," in which preschool children were offered the choice between eating one marshmallow right away or waiting a few minutes (actually up to 15 minutes) to receive TWO marshmallows (38a).

THE POLICY IMPLICATIONS OF HYPERBOLIC DISCOUNTING: THE CASE OF CIGARETTE TAXES

What are the implications of hyperbolic discounting as compared with exponential discounting? There are at least two policy implications of the behavioral economics model, and we use the example of cigarette smoking to illustrate the contrasts. In the standard economic model of cigarette smoking, fully informed, forward-looking, rational consumers make the decision to smoke after weighing the benefits of smoking (enjoyment) against the costs (premature mortality). The sole justification for taxation in this case is to recover the external costs imposed by smokers on the rest of society, for example, the cost to taxpayers of treating diseases caused by cigarette smoking.[7] However, by most economic calculations, these externalities are fairly modest (about 40¢ per pack), and it could be argued that smokers already pay their way through the taxes that are already levied on cigarettes—as a matter of fact, at current levels of taxation, smokers are most likely *subsidizing* the health care of nonsmokers (not to mention their social security). In short, one of the criticisms of the cigarette excise tax is that they are regressive, particularly since low-income individuals are much more likely to smoke compared with high-income individuals (39).

In contrast to the standard economic model of smoking, the behavioral economics model posits that smokers' preferences are time inconsistent. In this model, the smoker is temporally dispersed, that is, today's "self" gives greater weight to the immediate short-term pleasure of smoking while discounting future costs. Tomorrow's "self" is more patient, and would prefer to quit. But the problem is that "tomorrow" never comes. That is, confronted with the choice to quit now or quit tomorrow, the smoker falls prey to immediate gratification (stated another way, the value of deferring the nicotine hit till later is massively discounted). The same individual is much more patient when it comes to decisions in the distant future. For instance, he would be quite willing to entertain quitting in a month's time. The problem is that when we approach the same person in a month's time, he is back again to today's self. This model of intertemporal bargaining has been likened to a process of negotiation between temporally dispersed selves (the short-term "self" versus the long-term "self") that lead to behaviors that are ultimately against the interest of the unitary self.

The implication of this behavioral economics model of smoking is that our System 2 is often self-aware of our difficulties. Accordingly, System 2 may attempt to take action to incapacitate our future selves acting under the influence of System 1.[8] In practical terms, a hyperbolic discounter will seek to control their behavior by setting *commitment devices* to avoid temptation. Economists distinguish between two forms of commitment: *(1)* those taking the form of an excise tax (i.e., taxes as a self-control device demanded by time-inconsistent smokers who would like to quit

7 Some people point out that the premature mortality and diseases caused by smoking also result in lost productivity to society. However, in economic analysis, these "costs" are not treated as externalities; they are rather internal costs assumed by the smoker (and their families), which were presumably already factored into the decision to smoke.

8 In behavioral economics, reference is often made to Book XII of Homer's *Odyssey*, in which the hero lashes himself to the mast of his boat so that he can listen to the ravishing song of the Sirens and at the same time avoid the fatal side effect of diving into the waves and drowning.

but cannot), and *(2)* precommitment contracts entered into by the smoker as a form of betting against the self.

Regarding the former, behavioral economists predict that time-inconsistent consumers will have a demand for commitment devices that can be used to induce more desirable behavior in the present. Unfortunately, the private market only imperfectly provides self-control devices. Left to their own devices, smokers provide suboptimal commitment devices for themselves. But government can provide an excellent commitment device—in the guise of taxation. Assuming a modest degree of time inconsistency, the behavioral economics model suggests an optimal tax of between $4 and $14 per pack, that is, a much higher level of taxation than suggested by the cost-recovery model in the standard economic model (39). How did they arrive at the figure of $4–$14? To recap, the level of tax based on recovering the externality costs of smoking amounts to about 40 cents per pack. But the hyperbolic discounting model implies that we should also levy an *internality tax*, that is, the costs to the self, imposed by impatience and attendant overconsumption (40). Depending on the standard valuation of a life (as well as assumptions about the fraction of discounted health damages ignored by a hyperbolic discounter), the implied tax level comes out to between $4 and $14 (39). Stated another way, $14 is what your inner System 2 would suggest as the appropriate level of tax that reflects the internality cost imposed by your inner System 1.

The behavioral economics model of smoking also implies that smokers would be happier as a result of having an excise tax slapped on cigarettes. The standard economic model predicts that the smoker would be less happy as a result of having to pay an excise tax—after all, any tax on cigarettes treads on consumer sovereignty and smokers already pay more than their share of external costs induced by their habit. By contrast, the behavioral economic model predicts that if smokers have a latent demand for precommitment devices, they ought to be happier as a result of such a device being supplied by the state in the form of excise taxes. In an analysis of the General Social Surveys, 1973–1998, Gruber and Mullainathan (41) report that each $1 hike in cigarette taxes does seem to reduce the probability of unhappiness by 2.5 percentage points among individuals who have a propensity to smoke. Importantly, no such association is seen for paying other kinds of excise tax.

Realistically, there is a limit to how much tax can be levied on a product such as cigarettes. Smokers might be made happier but government treasurers might be less happy as a result of the unintended consequences of steep tax increases. A cautionary tale is provided by Canada in this regard. When Canada raised its federal cigarette tax on a 200-cigarette carton to $16 Canadian dollars in the 1990s, cross-border smuggling increased sharply along the St. Lawrence River. By the early 1990s, an estimated 30% or more of cigarettes smoked in Canada were smuggled in from the United States (80% of which were Canadian brands exported to the United States then smuggled back illegally). The illegal trade forced Canadian government to back down on its tax increase (42).

COMMITMENT DEVICES

As an alternative to raising the excise tax, precommitment devices can also take the form of personal *deposit contracts* to quit smoking. Ian Ayres (43) usefully distinguishes between

incentives and commitments. Whereas an incentive (such as a 40-cent excise tax on a packet of cigarettes) is intended to guide consumer choice, a commitment refers to some sort of device that *disables* choice. For instance, a deposit contract worth 6 months' salary (which the smoker agrees to forfeit should they fail to quit in half a year's time) would be a commitment device to incapacitate the future "self" acting under the influence of hyperbolic discounting.[9] It is the metaphorical equivalent of Odysseus lashing himself to the mast. Exactly this type of idea was implemented in a World Bank field study conducted on the island of Mindanao, Philippines (44). In this randomized controlled trial conducted by members of the Poverty Lab, the researchers approached two thousand smokers in the streets of Mindanao and offered them the opportunity to sign up for a commitment contract to stop smoking. The contract required each smoker to deposit money in a bank account (earning zero interest) for 6 months. At 6 months, if the smoker failed a urinary cotinine test, he forfeited his money (the money was donated to charity). The trial included two additional groups: *(1)* a group who received wallet-sized "cue cards" showing frightening pictures of the health consequences of smoking, and *(2)* a group who received no further inducements. The program thus provided two forms of voluntary commitment: a financial commitment in the form of savings balances, and a commitment to be visited by a deposit collector (and thereby receive the social pressure that may accompany such a visit).

Approached "out of the blue" on the streets of Mindanao, eleven percent of smokers took up the offer to sign up for the experiment. However, among those randomized to the treatment group, the average smoker made a deposit every 2 weeks and ended up committing 550 pesos (US $11) by the end of the 6-month contract period. Although $11 may not seem like a lot of money, it is equivalent to about 20 percent of the monthly income in Mindanao, or roughly equal to the average out-of-pocket expense for about 6 months' worth of cigarettes. In other words, the participants had a meaningful amount of "skin in the game." The results of the trial showed that smokers who were randomized to the deposit contracts were 2.8 to 5.7 absolute percentage points more likely (by intention-to-treat) to pass the 6-month urine test compared with the control groups. The program was by no means a panacea; the success rates for cessation at 6 months were still below 20% for all three groups—it was 18.1% in the deposit contract arm, 15.3% in the cue-card arm, and 12.4% in the do-nothing arm. Nevertheless, the study offers a tantalizing clue that the theory of precommitment boosts the success of behavior modification programs. How did the researchers know that the individuals in the treatment arm were not cheating, that is, stopping smoking just before the 6-month scheduled visit in order to wash out cotinine levels from their blood? The trial addressed this possibility by offering a random (i.e., a surprise) cotinine test at 12 months in a subset of subjects. About 60% of the subjects in the trial agreed to this component of the trial. And once again, those who took up the contract were 3.4 to 5.7 absolute percentage points more likely to pass the 12-month urine test compared to the two control groups.[10]

9 Similarly, a $14 tax hike on a pack of cigarettes goes well beyond an incentive—it would disable choice for a large number of smokers.

10 Along similar lines, a popular website on the Internet—called stickk.com—allows people to sign up for deposit contracts to commit to behavior change (Ayres, 2010). Its developers have called it a "commitment store" where people can sign up to commit themselves to reforming their habits. As of October 2013, over 200,000 individuals had signed up on the site, staking over $15 million of their own money to commit themselves to lose weight, stop smoking, stop chewing

SUPERCHARGING INCENTIVES

In designing any intervention to effect behavior change, Ayres (43) has noted that there are pros and cons of incentives versus commitments. Incentives (such as raising the price of a packet of cigarettes by 25 cents) can be revenue-generating, provided that the increase in revenue from the tax increase offsets the drop in demand. By contrast, commitment contracts can be expensive if they take the form of a large reward to perform a behavior; for example, if an employer decides to offer several thousand dollars to any worker who quits smoking. At best, commitment contracts will end up being revenue-neutral, if smokers can be persuaded to commit their own money to quitting contracts.

Which is the better way to motivate behavior change—offering monetary rewards as an incentive, or asking individuals to bet their own money on a deposit contract? In other words, should we use the carrot or the stick? To answer this question, Volpp and colleagues (45) carried out a worksite-based weight loss intervention. In their study, 57 healthy participants aged 30–70 years with a BMI of 30–40 were randomized to three weight loss plans: *(1)* monthly weigh-ins, *(2)* a lottery incentive program (*the carrot*), or *(3)* a deposit contract (*the stick*) with a weight loss goal of 1 lb (0.45 kg) a week for 16 weeks. The intervention deliberately embeds several principles from behavioral economics in order to "supercharge" people's motivation, that is, the principle of precommitment (people are willing to commit their future selves to some course of action, even though they would rather not have to do it now), loss aversion (people hate to lose something they have steadily accumulated, such as a deposit), and people's tendency to overweight small possibilities (such as winning a lottery). We shall describe loss aversion more formally in the next section.

In the group randomized to the lottery incentive, participants were eligible for a daily lottery prize if their weight was at or below their goal. The lottery provided frequent small payoffs (1 in 5 chance of winning $10) as well as infrequent large payoffs (1 in 100 chance of winning $100). The results of the daily lotteries were texted directly to the participants in order to provide instant, tangible feedback. The daily texting of lottery results leverages people's present-focused preferences—described as a decision "error" in the previous sections, but in this instance, turned into an actual advantage—by making the rewards for healthy behavior seem immediate and salient—as opposed to rewarding them with extra compensation buried somewhere in their monthly paycheck. In the deposit contract arm, the subjects were given the opportunity to deposit between .01 and $3.00 for each day of the month. As an added incentive, the employer matched the employee contributions 1:1, and added a fixed payment of $3 per day. At the monthly weigh-in, if the participants met their target weight, they received the accumulated money (up to a maximum of $252 per month). If the subjects failed to meet their target, their deposits would be forfeited and added to an office pool of money that was divided equally among deposit contract participants who lost 20 pounds or more over the 16 weeks. The behavioral economics principle behind this feature is precommitment as well as loss aversion (whereby the prospect of losing something you already have is much more aversive than winning something you don't have—see next section).

their fingernails, finishing their dissertations, and so on. An added twist to these personal deposit contracts is that people can pledge to give away their deposits to an "anticharity" of their choice, should they fail to accomplish their goals. For example, a pro-choice person can pledge in advance that their forfeited money can be donated to a pro-life lobby group, or a gun control advocate can pledge that their money can be donated to the National Rifle Association, and so on.

In the end, 90% of participants made deposits averaging $1.56 per day. Comparing the two active arms of the trial over four months, those in the deposit group earned on average $378, while the lottery group earned on average $273. At the end of the 4-month trial, the incentive groups lost significantly more weight than the control group (mean, 3.9 lb). Compared with the control group, the lottery group lost a mean of 13.1 lb (95% confidence interval [CI] of the difference in means, 1.95–16.40; p = 0.02) while the deposit contract group lost a mean of 14.0 lb (95% CI of the difference in means, 3.69–16.43; p = 0.006). In other words, when the intervention was supercharged with behavioral economics principles, the carrot (lottery incentive) and the stick (deposit contract) proved to be equally effective. About half of those in both incentive groups met the 16-lb target weight loss: 47.4% (95% CI, 24.5%–71.1%) in the deposit contract group versus 52.6% (95% CI, 28.9%–75.6%) in the lottery group, whereas 10.5% (95% CI, 1.3%–33.1%; p = 0.01) in the control group met the 16-lb target.

Despite success over a 4-month period, however, behavioral maintenance remains a huge challenge. Behavioral interventions that leverage insights from behavioral economics have overwhelmingly focused on short-term behavior change, or even single-time behaviors (such as getting a flu shot). The long-term maintenance of behavior change still remains the Holy Grail of interventionists—a challenge that is shared by virtually all models of behavior change. In the study by Volpp and colleagues, when follow-up was extended to 7 months (i.e., three months after the intervention had stopped), both intervention groups had regained between a third to a half of the weight. Although the net weight loss (relative to baseline) at the end of 7 months was larger in the incentive groups (9.2 lb; t = 1.21; 95% CI, -3.20 to 12.66; p = 0.23, in the lottery group and 6.2 lb; t = 0.52; 95% CI, -5.17 to 8.75; p = 0.61 in the deposit contract group) compared with the control group (4.4 lb), the differences were not statistically significant.

Two additional questions concerning the design of optimal incentives seem germane to discuss here. First, it is important to get the size of the incentive just right. Too small an incentive will fail to motivate, while too large an incentive can backfire (for example, by killing intrinsic motivation). A second randomized trial conducted by Volpp et al. (46)—this time, to motivate smoking cessation—addressed this issue. In this trial, 878 employees of a firm were randomized to either an education intervention (provision of information about a smoking cessation program), or information combined with financial incentives. The incentive structure was as follows: each employee received $100 for completing the smoking cessation program, another $250 for biochemically confirmed cessation at 6 months, plus $400 more for confirmed cessation at 12 months, for a maximum $750 reward. At 9–12 months follow-up, the incentive group had significantly higher rates of smoking cessation than did the information-only group (14.7% vs. 5.0%, p < 0.001). At 15–18 months follow-up, both groups experienced some relapse; nonetheless, the incentive group still showed higher cessation rates compared to the information-only group (9.4% vs. 3.6%, p < 0.001). The results of this intervention are noteworthy because according to a Cochrane Collaboration systematic review of incentive-based interventions for smoking cessation, it was the only trial out of nineteen found in the literature to demonstrate a statistically significant increase in quit rate beyond 6 months (47). One reason is that the study by Volpp et al. (46) was the largest study conducted (representing almost 20% of the total patient sample size included in the meta-analysis) and more likely to have adequate power to detect an effect. An additional possibility is that the size of incentives used across studies was generally small—much too small, and in some cases as low as $10 to quit smoking (47).

In short, if incentive-based interventions are to be tried, it is critical to get the size of the incentive right. The second issue to bring up regarding the use of incentives is that they need to be matched to the stage of preparation (a construct from the transtheoretical model of change [48]). In the trial by Volpp and colleagues (46), subgroup analysis suggested that the incentive was effective in boosting smoking cessation only among smokers in the contemplation and preparation stages of change. Among precontemplators, the incentive did not work. Furthermore, Kim et al. (49) examined the differences in the level of monetary rewards that smokers said they would need to be paid in order to be incentivized to quit, according to their stages of change. In line with theory, 42% of smokers in the preparation stage said that they would quit for a reward between $1 and $750, whereas 46% of those in the precontemplation stage said they needed $3,400 or more.

The suggestion that for some people a very large incentive might be needed to motivate behavior inevitably brings up the question: *Who pays?* In a healthcare system such as the United States with a substantial private insurance component, the answer needs to take into account *who benefits* from incentivizing healthy behavior (50). For instance, in a high turnover private insurance market, it would not make a lot of sense for the insurer to pay out a ton of money to incentivize people to attempt to change behaviors with a long-term payoff because the cost-savings would end up being captured by someone else. Hence in high turnover markets, it may be more cost-effective to focus on incentives with short-term payoffs—for example, asthma management or smoking cessation. By contrast, in the case of long-term insurance providers (Medicare, VA), it may be cost-effective to focus on long-term payoffs such as improved blood pressure control or weight loss. Finally, for problems with a substantial public health interest (for example, adherence to medication among patients with multidrug resistant tuberculosis), the state needs to step in to counteract public externalities (50).

FRAMING EFFECTS AND LOSS AVERSION

Along with heuristic biases, default options, and dynamically inconsistent preferences, one of the most significant discoveries in behavioral economics is the demonstration of framing effects and loss aversion. The *framing effect* refers to a cognitive bias in which people react differently to a choice depending on whether it is presented as a loss or as a gain. In the original example described by Kahneman and Tversky (51) in their *Econometrica* paper, people switched preferences depending on the way in which the choices were framed. I have tweaked the scenario slightly to make it more vivid. Suppose the CDC announces that a new and deadly zoonotic virus has erupted and is threatening the country. If nothing is done, the virus will kill 600 people. They will die a horrible death by "bleeding out" into their internal organs and collapsing in a puddle of blood. Fortunately, two alternative treatments have been developed:

- If Serum A is adopted, 200 people will be saved;
- If Serum B is adopted, there is:
  - 1/3 probability that 600 lives will be saved; and
  - 2/3 probability that nobody will be saved.

Which serum would you choose? In the above scenario, more people tend to choose serum A than serum B. How about a different scenario involving a choice between two other treatments, serum C and serum D?

- If Serum C is adopted, 400 people will die.
- If Serum D is adopted, there is:
  - 1/3 probability that nobody will die; and
  - 2/3 probability that 600 people will die.

In this scenario, more people seem to favor serum D. What happened? When we look at the expectancies under each scenario, serum A offers the same odds as serum C; both of them will result in 400 people dying. The only difference is that serum A is framed in terms of the certainty of survival ("200 people will be saved"), while serum C is framed in terms of the certainty of death ("400 people will die"). Accordingly, a person choosing serum A > serum B in the first choice set ought to also choose serum C > serum D in the second choice set. Yet that is not what happens. When the prospect of certain gain is presented to people (as in serum A), they tend to become risk-averse, and prefer to avoid the gamble represented by serum B. By contrast when the prospect of certain loss is dangled before people (as in serum C), people tend to become risk-seeking, preferring to take a gamble by opting for serum D. Clearly, this represents an "irrational" preference switch, since serum A and C are equivalent (and so are serum B and D). The only thing that changed was how the choice was described (framed)—an irrelevant feature of the choice that ought not to distract the rational decision-maker. Nonetheless, the tendency for people to prefer a certain gain over a certain loss seems to be a fairly robust feature of decision-making that Kahneman and Tversky called loss aversion.

The preference for desiring a "bird in the hand" (which is proverbially worth two in the bush) seems to be an evolutionarily preserved feature of primate behavior. For example, in experiments involving caged capuchin monkeys, Chen et al. (52) trained the animals so that they expected to receive slices of apple when they moved to one side of the cage or the other. The two sides of the cage either displayed one or two slices of apple. On the side of the cage that displayed two slices of apple, when the monkey approached this side, 50% of the time he received both slices, but the other 50% of the time the experimenter removed one slice just before the monkey reached the apple slices, and consequently the monkey only ended up with a single slice. On the other side of the cage, there was always one slice of apple. When the monkey approached this side, 50% of the time the experimenter doubled the slices (i.e., the capuchin received two slices), and the remaining time the monkey would just get the one slice. To sum, the two sides of the cage offer equivalent gambles, that is, a 50:50 chance of either ending up with one slice or two slices. A hyperrational monkey ought to be indifferent to which side he prefers to go to. But the capuchins in this experiment overwhelmingly preferred the side with one slice (70%) as opposed to the side displaying two slices. Evidently, the capuchin prefers the sure prospect of obtaining one slice as opposed to the 50% chance of losing one slice.

Just as in the case of hyperbolic discounting, the phenomenon of loss aversion also has a basis in neuroscience. For example, De Martino et al. (53) put subjects through a set of choice tasks that were deliberately designed to mobilize loss aversion. Each subject received a £50 honorarium. Then, while they were being scanned in the fMRI machine, they were asked which they would

choose: *(1) keeping £20* of their honorarium money for certain, or *(2)* taking a gamble with a 2/3 probability they would lose all the money, and a 1/3 chance they would keep all the money (i.e., £50). This is a *gain-frame*, which should drive more people to become risk-averse, and that is exactly what they found—53% preferred to keep £20 for sure, rather than risking a gamble. In the alternative condition, people were asked which they would choose: *(1) losing £30* of their honorarium, or *(2)* taking a gamble with a 2/3 probability they would lose all the money, and a 1/3 chance they would keep all the money (i.e., £50). This is a *loss-frame*, which should drive more people to become risk-seeking, and that is exactly what they found—62% preferred to gamble, rather than risk the certainty of losing £30. (Of course, by now the reader should have cottoned on to the fact that in this scenario, "keeping £20 out of the original 50" is the same as "losing £30 out of the original 50," and hence it is "irrational" to prefer one over the other). The fMRI scan results showed that the framing effect was specifically associated with amygdala activity, suggesting a key role for the emotional system (System 1) in mediating decision biases. In other words, the part of the brain associated with fear and anxiety seems to put the cautionary brake on our behavior that we call loss aversion. Conversely, across individuals, higher activity in the orbital and medial prefrontal cortex activity (System 2 areas) predicted a reduced susceptibility to the framing effect. The results suggest an opponency between the two neural systems.

APPLICATION OF FRAMING EFFECTS TO HEALTH MESSAGES

One of the earliest applications of the principle of loss aversion to a behavior change intervention was described by Banks et al. (54), in which the researchers hypothesized that the framing of a message about the benefits of cancer screening (mammography) would influence uptake. Specifically, they conjectured that women view getting a mammogram as a risky procedure (since they "risk" having a tumor detected). Therefore, in order to induce them to be risk-seeking (i.e., come forward for a mammogram), messages about mammograms would be more effective by being loss-framed—for example, by evoking the loss of life because of undetected cancer. One hundred and thirty-three women 40 years and older and not adhering to current guidelines for obtaining mammography screening were assigned randomly to view either gain-framed (emphasizing the benefits of obtaining mammography) or loss-framed (emphasizing the risks of not obtaining mammography) persuasive videos that were factually equivalent. Mammography utilization was assessed 6 and 12 months later. Consistent with predictions based on loss aversion, women who viewed the loss-framed message were more likely to have obtained a mammogram within 12 months of the intervention (66% vs. 51%).

In contrast to cancer screening (which the researchers viewed as a "risk-seeking" behavior), other types of preventive behaviors—such as applying sunscreen—can be viewed as essentially riskless. According to theory, then, gain-framed messaging ought to be more persuasive than loss-framed messaging. Detweiler and colleagues (55) put this idea to the test among 217 beach-goers. Attitudes and intentions about the use of sunscreen were measured before and immediately following the delivery of gain- and loss-framed informational brochures, and after

completing the questionnaire participants were given a coupon redeemable for a small bottle of sunscreen later that same day. People who received the gain-framed brochures, compared with those who received the loss-framed brochures, were significantly more likely to redeem their coupons for sunscreen while on the beach (71% vs. 53%).

How robust are these findings? In a systematic review, O'Keefe and Jensen (56) meta-analyzed the 93 published studies (involving 21,656 subjects) that compared the differential efficacy of loss-framed versus gain-framed messages for inducing behavior change. They found that for riskless disease prevention behaviors, there was a statistically significant but tiny ($p = 0.03$) advantage of gain-framed appeals over loss-framed appeals. On the other hand, they found no statistically significant difference in persuasiveness comparing gain-framed versus loss-framed messages concerning other preventive actions such as safe-sex behaviors, skin cancer prevention behaviors, or diet and nutrition behaviors. They concluded: "Sufficient research evidence has accumulated to make it plain that the use of loss-framed rather than gain-framed appeals will generally not make much difference to persuasive success in [improving uptake of cancer screening behaviors]" (56).

With hindsight it appears that the initial application of prospect theory (of which the concept of loss aversion is a part) to design health messages might have been a bit too crude. The situation is more complex than assuming that the framing of cancer as a "loss" would encourage more women to take a gamble on getting screened. For this assumes that women construe cancer screening as a "risky" decision, which may not be the case (57). Indeed one study suggested that a loss-framed brochure for breast self-examination only worked among women who perceived the procedure as risky (58). In short, there are other features of the screening decision that may be equally important (or perhaps more important) than how the message is framed. For instance, the affect heuristic—invoking negative emotions such as fear in the case of loss-framed messages, or positive emotions such as hope in the case of gain-framed messages—may be just as persuasive in motivating a decision. In addition, most forms of cancer screening involve the problem of intertemporal choice discussed in the previous section, that is, the subject has to be persuaded to pay a cost upfront (in the form of anxiety, taking time off work, sometimes pain or discomfort) in order to avert a bad outcome later. Given this problem, framing the message in a particular way may not suffice to get people screened; they may develop good intentions to get screened, but never get around to actually following through with it because of procrastination. In that case, it may be necessary to incorporate additional tactics to motivate behavior, such as the use of precommitment devices (for example, Dan Ariely (59) suggests that health plans might ask employees to voluntarily pay a $300 deposit for a colonoscopy when you reach age 50; if you comply, you get your money back).

FRAMING EFFECTS IN MOTIVATING HEALTHY EATING

Notwithstanding the disappointing results from the application of framing effects to cancer screening messages, there is still plenty that could be done to leverage this effect in other areas of health promotion. Framing effects in real life are ubiquitous but underutilized. In an amusing

experiment conducted in an MIT classroom, Dan Ariely and colleagues (60) randomly allocated students to one of two conditions; both involved Professor Ariely reading aloud poetry to the class from *Leaves of Grass* (when presumably, the time ought to have been spent learning about behavioral economics). For half the class, the poetry recitation was framed positively, that is the students were asked whether they would be *willing to pay* $10 to listen to the professor recite poetry, and if so, how much they would be willing to pay to listen to him read for 1, 3, and 6 minutes. In the remaining half of the class, the task was framed negatively, that is, the students were asked whether *they would accept compensation* of $10 to put up with listening to the professor recite poetry; and if so, how much money they would need to be paid in order to listen to him for 1, 3, and 6 minutes. The results showed that students' valuations were strongly influenced by the initial frame. Students who were given the positive frame were, on average, willing to pay for the experience, while students in the negative frame, on average, demanded compensation for undergoing the same experience. Moreover, respondents consistently indicated higher sums of money for longer durations, regardless of whether they happened to be in the payment mode or the compensation mode. As Ariely discovered, the students did not have any prior sense of whether the poetry recital was good or bad, worthwhile or worthless; but they all "knew" that being exposed to more of the experience warranted greater payment.

One implication of these findings from behavioral economics is that if we want to motivate people to perform a health behavior (say, eating more vegetables), then we need to pay much more attention to how we frame the message—for example, should our message focus on the nutritional benefits of the tomato ("good source of lycopene"), or focus on the delectable taste of the Brandywine heirloom tomato in season? Of course we should do both—that is, promote the nutritional value of vegetables and their flavor. Nonetheless, caution is also warranted in the way we use health claims to frame the message. This was demonstrated in a series of experiments by Finkelstein and Fishbach (61). They found that when snacks were framed as "healthy" (as opposed to "tasty"), people reported feeling hungrier afterward, and ended up consuming more calories to compensate. In the experiment conducted on a mid-Western college campus, students were offered a tasting of a slice of bread. In the healthy frame, the bread was described as "nutritious, low-fat, full of vitamins," while in the tasty frame, the same slice of bread was described as "yummy with a thick crust and soft center" (and no mention of nutritional benefits). After the tasting, the participants were shown to a second room in which they filled out a survey. On a table, the researchers left a bowl of pretzels to which subjects could help themselves as they completed the survey. The result was that people who tried the "healthy" slice reported feeling more hungry afterward than people who tried the "tasty" bread; the former group also helped themselves to significantly more pretzels, and the difference was most marked among people who said they were not watching their weight.

In a separate experiment, the researchers examined the effect of forcing people to eat healthy (imposed consumption). In this version, people were shown two protein bars, one of which was labeled "healthy" and the other which was labeled "tasty" (in fact, they were identical). In the imposed consumption condition, the participant was directed to sample either the healthy or tasty bar based on random assignment, that is, even if the subject preferred to eat the tasty bar, they were told to try the healthy one). In the free choice condition the participant could pick the one they liked. The results showed that among subjects in the imposed consumption condition, those who were forced to eat healthy caused a rebound in their short-term experience in

hunger—their rating of hunger afterward was higher than any other group. To summarize these findings, promoting better nutrition through appeals to "healthy" can backfire. In public health there is a strong tendency to assume that the appeal to "healthy" ought to be irresistible to people. But these experiments suggest that sometimes framing a food as "healthy" can result in increased feelings of hunger and rebound caloric intake.[11]

Thus, knowledge of the framing effect can help to not only design better messages, but also to understand which traps to avoid. One last example of the unintended side effect of "healthy" labeling stems from an experiment conducted in a lunch-hour restaurant in New Haven (62). In this experiment the researchers convinced the restaurant owners to alternate between: *(1)* price reductions, *(2)* health messaging, and *(3)* a combination of the two on the purchase of healthy food items on the lunch menu. For example, during the weeks of the price reduction intervention, a 20–30% price cut was offered on the low-fat grilled chicken sandwich. During the weeks of the health messaging campaign, the store displayed prominent messages such as "Healthy eating increases physical & mental well-being" next to the healthy items. In terms of the impact of these strategies, the price decrease beat health messaging handily; there was far more customer response to the price cut than to the health appeals. What happened when the two strategies were combined? During the weeks when customers were offered both the price cut and the health message, the sales of the healthy items did go up, *but not by as much as* when just the price cut was offered. In other words, adding the health message to the price cut messed up the effect of the discount. One interpretation of this result is that when customers see the price of a "healthy food" being knocked down, they conclude that "it must taste bad, and that's why they are selling it at a discount." Indeed by now there seems to be a general image in the public's mind that "healthy" equates with "tastes bad" (which is one reason why Miller Lite commercials take such pains to emphasize its "great taste…less filling" qualities). It follows that if we want to encourage more consumption of broccoli among the public (or our kids at home), the last thing we should be touting is its nutritional benefits. A further corollary is that the framing of foods as "healthy" is only really useful to people who are trying to sell *unhealthy* food. Hence the world of food marketing is replete with examples of spurious health-based claims. These claims are probably effective because the shopper already knows that the product they are reaching for in the supermarket aisle is bad for you, but the prominent health claim on the front of the package ("full of vitamins!" "organic!" "zero trans!") temporarily soothes or disables the monitoring function of System 2 over System 1.

CONTROVERSIES IN BEHAVIORAL ECONOMICS APPLICATIONS TO POLICY

In certain policy circles, the concept of "nudge" has found traction because it seems to offer a hybrid (and possibly oxymoronic) philosophy of state intervention that Thaler and Sunstein (63)

11 As a fascinating side note, the authors of this *Journal of Consumer Research* article drew an entirely different implication of their study results, as follows: "Supermarket owners might consider offering more 'healthy' food samples in order to promote more in-store sales." (!)

have dubbed "libertarian paternalism." As defined by Thaler and Sunstein, libertarian paternalism is defined by policies that "shift behavior in self-interested (as opposed to self-destructive) directions *without abridging the individuals' ultimate freedom to choose*." For example, in the earlier example of the sandwich restaurant menu, the healthy default option was to display the low-calorie menu items at the front. This nudges the customer in the direction of ordering those items, but if they really insist on ordering the high-calorie hoagie, they can flip to the back of the menu and still find those items. In other words, nobody is banning these foods. Ditto, in Mayor Bloomberg's proposed restriction on sales of sodas over 16 ounces, if someone really has a hankering for drinking that amount of Mountain Dew, they could still purchase two 8-ounce beverages and get the same effect.

According to Camerer's (64) typology of policy interventions, the "nudge" falls under the category of *asymmetric paternalism*. When we consider who benefits and who pays the price of various regulations, we can see that some regulations are redistributive, that is, they benefit some people at the expense of others—for example, taxing the rich to provide for the poor. In other cases, the regulation is intended to prevent the individual from harming others (counteracting externalities)—such as indoor smoking restrictions. Yet other regulations force (or prevent) choices for the individual's own good, that is, hard paternalism—which is the case for prohibition of crack cocaine. The problem with paternalism (from a libertarian's point of view) is that it treads on consumer sovereignty. However, not all forms of paternalism are created the same. Asymmetric paternalism refers to when a policy creates large benefits for individuals who are *boundedly* rational, while imposing little or no harm for those who are fully rational. In other words, for those of us who have self-control problems, the nudge can be quite beneficial in steering our behavior toward healthy directions. For others who have rationally picked the hoagie as their lunch of choice (i.e., they have weighed the costs and benefits), they are available on the menu.

So much for the philosophy. The problem is that when behavioral economics is translated to the realm of politics, it can morph into something else. Alas, the story is not at all unfamiliar. Whenever there is an enthusiastic rush to apply a new idea to practice, it runs the risk of "capture" by those who wish to use it to advance their own agenda. Such seems to have been the story in Britain, where the Prime Minister's Office established a Behavioral Insights Team (advised by Richard Thaler) to implement behavioral economics ideas into public policy. The problem seems to be that nudges were perceived as being sold as a *replacement* for existing (and more effective) health regulations. We can get the gist of the objections to the British government's nudge strategy from Rayner and Lang (65):

> The UK coalition government's enthusiasm for nudge . . . dispenses with the complexity of real life contexts and acknowledges only the immediate proximal horizons of consumer choice. At a stroke, policy is reduced to a combination of cognitive and "light" environmental signals, such as location of foods within retail geography. Nudge, along with the responsibility deals [with the food and alcohol industry] is presented as the alternative to regulation, or in the media jargon, the "nanny state." Our worry is that nudge becomes collusion between the state and corporations to hoodwink consumers. At least nannies are overt.

The controversy eventually erupted in the pages of British medical journals—including the *Lancet* and *BMJ*—during 2011. Just reading the titles of these opinion pieces provides a glimpse of the heated nature of the debate: "One Nudge Forward, Two Steps Back" (66); "Nudge Smudge: UK Government Misrepresents 'Nudge'" (67); "Judging Nudging: Can Nudging Improve Population Health?" (68).

There is cogency to the argument of the critics. Behavioral economics cannot be used with the aim of replacing or dispensing with effective but more paternalistic interventions such as excise taxes on cigarettes. Behavioral economics is not a substitute for established approaches to health promotion, including taxation and regulation. George Loewenstein and Peter Ubel (69) sounded the same warning in an op-ed piece in the *New York Times*, which is worth quoting at length here:

> It seems that every week a new book or major newspaper article appears showing that irrational decision-making helped cause the housing bubble or the rise in health care costs. Such insights draw on behavioral economics, an increasingly popular field that incorporates elements from psychology to explain why people make seemingly irrational decisions, at least according to traditional economic theory and its emphasis on rational choice. Behavioral economics helps to explain why, for example, people under-save for retirement, why they eat too much and exercise too little…
>
> "But the field has its limits. As policymakers use it to devise programs, it's becoming clear that behavioral economics is being asked to solve problems it wasn't meant to address. Indeed, it seems in some cases that behavioral economics is being used as a political expedient, allowing policymakers to avoid painful but more effective solutions rooted in traditional economics.…
>
> Behavioral economics should complement, not substitute for, more substantive economic interventions. If traditional economics suggests that we should have a larger price difference between sugar-free and sugared drinks, behavioral economics could suggest whether consumers would respond better to a subsidy on unsweetened drinks or a tax on sugary drinks. But that's the most it can do. For all of its insights, behavioral economics alone is not a viable alternative to the kinds of far-reaching policies we need to tackle our nation's challenges.

The relevance of behavioral economics for public policy will only increase over time. In the era of spiraling healthcare costs, calls for more "patient engagement" have become the buzzword. In a recent survey of employers, 61 percent of employers identified "employees' poor health habits" as the number one challenge for maintaining affordable benefits (70). Whether we agree or not with employers' diagnosis of their problems, the fact remains that paying for the performance of patients (P4P4P) is an idea that is gaining increasing traction. Indeed in the Affordable Care Act ("Obamacare"), Section 2705 stipulates that beginning in 2014, employers may use up to 50% of the total amount of employees' health insurance premiums to provide outcome-based wellness incentives. These incentives can take the form of "discount/rebate on premiums, or waiver of cost-sharing mechanisms (deductibles, co-payments, co-insurance)" (50). In other words, public health needs to be better prepared to provide the necessary evidence base when insurance companies come knocking.

BEHAVIORAL ECONOMICS AND SOCIAL INEQUALITIES IN HEALTH

To summarize the state of knowledge, we can say that the effect sizes found in behavioral economics-based approaches to behavior modification are still rather modest. Behavioral economics can boost the success of traditional health promotion programs, but the field is still in its infancy (at least as far as being translated to scalable policies) and thus far from rising to Len's Grand Challenge introduced at the beginning of this chapter. In one sense, the "low-hanging fruit" has already been plucked—for example, whopping excise taxes are already levied on cigarettes and regulations have been expanded to restrict smoking in public places (and lately, even to low-income public housing [71]). We are left with the "last mile problem," namely, how to encourage behavior change in the remaining, resistant segments of the population who contribute to the persistence of these problem behaviors, as well as the lingering socioeconomic disparities in health behaviors. To say that we need to improve the living conditions of disadvantaged groups is tantamount to a restatement of the "last mile problem;" it is going to be a long, hard slog.

By acknowledging the "nonrational" influences on behavior—emotion-based decision-making, inconsistent time preferences, framing effects—I believe that behavioral economics offers additional avenues for addressing socioeconomic inequalities in health behavior. As almost everybody acknowledges, the behavior modification programs of the past have proved immensely successful in persuading high-resource individuals to alter their behavior. Stated another way, the success of our past education efforts have partly contributed to the widening of the socioeconomic gaps. Highly educated individuals pay more attention to health education messages, and they have the cognitive resources to translate those messages into action. By contrast, low-SES individuals often lead chaotic lives, they are surrounded by a sea of bad choices, and in some contexts they may even have rational reasons to behave "badly." In the language of Mullainathan and Shafir (4), material scarcity imposes a bandwidth tax on our ability to plan for the future. At their best, the insights from behavioral economics can leverage people's decision errors and turn them to their advantage—they can be a kind of mental jujitsu. For example, using health communication strategies that do not rely exclusively on System 2 cognitions could help *everyone*, including the less educated, as well as the young. Using default options to guide behavior could nudge the whole population toward behaving more healthily, including people who are too preoccupied to pause and perform a cost-benefit calculus before making a decision.

As pointed out by Loewenstein et al. (70), existing programs being implemented by US employers, insurers, and healthcare providers already use incentives to encourage patients to take better care of themselves. But they are unlikely to have much impact because they "require information, expertise, and self-control that few patients possess" (70). Consequently, these programs disproportionately benefit those who are already taking good care of their health. To give one example, there is much talk of connecting health insurance premiums to health behaviors. This can take the form of raising premiums on smokers or obese workers (or conversely, lowering premiums on those who successfully quit or lose weight). The risk inherent in this approach is that if the incentives are not structured optimally, they may end up simply raising the premiums of those with health problems while leaving their behavior unchanged—a regressive result that simply punishes the most disadvantaged and vulnerable segments of the workforce. Given this risk,

employers could do better by incorporating insights from behavioral economics to design an optimal incentive structure. For example, in the previous sections, the use of behavioral insights to "supercharge" incentives was described (72). If the employer is prepared to pay workers to improve their behavior, why not do it properly? For example, instead of offering a reduction in insurance premiums for good behavior (which employees only notice—if at all—when the savings are folded into their monthly pay stub), why not take advantage of decision "errors" such as present-focused preferences to offer them smaller and frequent rewards that are more salient?

More recently, insights from behavioral economics have begun to be directly applied to the consideration of motivating change among low-SES populations. Richards and Sindelar (73) have suggested a set of policy proposals involving changes to the Supplementary Nutrition Assistance Program (SNAP). This is an instance of the application of behavioral economics principles to directly target the healthy food choices of low-income families. The authors first critique existing proposals to improve food choices in the SNAP program. For example, improving nutritional education to SNAP recipients is frequently cited. The problem—as we have seen throughout this chapter—is that there is limited evidence in support of the idea that the provision of information improves food choices. Another policy idea is to restrict the range of SNAP-eligible items, such as already happens in the Women, Infants, and Children (WIC) Program. The problem here is that SNAP represents a relatively small fraction of the monthly food budget of even low-income families (the average monthly benefit in 2012 was $130 per individual), and hence households may simply substitute their pattern of expenditure (i.e., use their personal income to continue to purchase unhealthy items). Yet a third proposal is to link the purchase of healthy food items to a cash-back subsidy, such as in a Massachusetts SNAP pilot. Although appealing in the sense that a food subsidy is less punitive and paternalistic than restricting the list of SNAP-eligible food items, there are formidable challenges in implementing a subsidy program, such as the costs of administering the program (e.g., keeping track of the subsidized food items, keeping receipts, and sending out the subsidy checks). In contrast to these existing proposals, Richards and Sindelar (73) propose novel interventions based on behavioral principles. One idea is to reward participants often and frequently with raffles for prizes if the participants meet a specified number of qualified purchases. The prizes could be home-cooking equipment or exercise equipment or any other kind of voucher for goods. Gneezy and List (74) explain the appeal of lotteries as a behavioral motivator, namely, people tend to overestimate the probability of rare events (such as winning the lottery), and this judgment error makes them that much more motivated to reach a goal. The same principle has been used effectively to improve academic performance in inner-city schools—see Gneezy and List (74).

A second proposal by Richards and Sindelar (73) leverages the idea of precommitment. Participants would first select their desired purchases from a comprehensive list of approved healthy foods, thereby committing themselves to higher nutritional quality ahead of time. This has the merit of bypassing the temptation caused by in-store cues as well as impulse purchases.[12] The program could even assist participants by offering sample shopping lists as a default. Participants

12 This is somewhat reminiscent of dinner meetings held at the Harvard Faculty Club. The dining room usually requires attendees to preorder their entrees ahead of time. Many a time, I have avoided falling prey to the "heat of the moment" by preordering the roasted halibut instead of the New York strip steak. More than once, however, I have observed that when the diners are served, the kitchen ran out of steak, and they were left with a surplus of halibut. It seems Harvard professors are not immune to inconsistent time preferences.

could always opt out of these defaults, but those who are preoccupied with other matters can just go with the defaults. Whether these policy proposals actually get taken up remains to be seen, but they illustrate how behavioral economics principles are actively being applied to tackle socioeconomic inequalities in health.

Lately, concepts from behavioral economics have even begun to be proposed in the setting of global health. Taylor and Buttenheim (75) discuss their application in the context of antiretroviral programs to prevent the transmission of HIV from mothers to their children (PMTCT programs). These programs are only effective to the extent that patients adhere to the medication regimens. Among suggested innovations, the authors cite the use of economic incentives, such as cell phone minutes or food vouchers that reward adherence to PMTCT protocols; as well as commitment contracts that incentivize mothers to return to the clinic in order to avoid both reputational and financial loss.

CONCLUSION

The future of behavioral science calls for an integrative approach that avoids the "contentious dualism" that pits individualist versus structuralist approaches to health behavior (see Chapter 10). Likewise, it is a false dichotomy to pit behavioral economics approaches against established models of behavior change. Indeed Zimmerman (76) has called for the development of a multi-level theoretical paradigm that acknowledges how our behavioral choices are shaped by heuristics and habits, but also simultaneously shaped by power imbalances in society that reflect social, political, and economic structures.

Insights from the field of behavioral economics are unlikely to—nor intended to—replace or compete with established models of behavior change. The emerging research from the fields of economics, psychology, and neuroscience is best viewed as complementing the established principles of behavioral science. Research is urgently needed to better integrate the emerging insights into multilevel behavioral interventions. Behavioral science (of whatever stripe) has still some way to go toward meeting Len Syme's challenge introduced at the beginning of the chapter, especially with regard to solving the problem of long-term behavioral maintenance. Progress is likely to be achieved only by building connections across complementary approaches.

REFERENCES

1. Keeney R. Personal decisions are the leading cause of death. Oper Res. 2008;56(6):1335–47.
2. Walker R, Block J, Kawachi I. Do residents of food deserts express different food buying preferences compared to residents of food oases? A mixed-methods analysis. Int J Behav Nutr and Phys Act. 2012;9(1):41.
3. Larson N, Story M. A review of environmental influences on food choices. Ann Behav Med. 2009;38(Suppl 1): S56–73.
4. Mullainathan S, Shafir E. Scarcity. Why having too little means so much. New York: Times Books; 2013.

5. Knowles J. The responsibility of the individual. In: Knowles J, editor. Doing better and feeling worse: health in the United States. New York: Norton 1977. pp. 57–80.
6. Thorgeirsson T, Kawachi I. Behavioral economics: merging psychology and economics for lifestyle interventions. Am J Prev Med. 2013;44(2):185–9.
7. Thaler R, Sunstein C. Nudge: improving decisions about health, wealth, and happiness. New York: Penguin Books; 2008.
8. Rose G, Marmot M. Social class and coronary heart disease. Br Heart J. 1981;45(1):13–9.
9. Marmot M. Status syndrome. London: Bloomsbury; 2004.
10. Stringhini S, Sabia S, Shipley M, Brunner E, Nabi H, Kivimäki M, et al. Association of socioeconomic position with health behaviors and mortality. JAMA. 2010;303(12):1159–66.
11. Syme S. Social determinants of health: the community as an empowered partner. Prev Chronic Dis. 2004; 1(1):A02.
12. Pennant M, Davenport C, Bayliss S, Greenheld W, Marshall T, Hyde C. Community programs for the prevention of cardiovascular disease: a systematic review. Am J Epidemiol. 2010;172(5):501–16.
13. King D, Mainous A 3rd, Carnemolla M, Everett C. Adherence to healthy lifestyle habits in US adults, 1988–2006. Am J Med. 2009;122(6):528–34.
14. Sorensen G, Himmelstein J, Hunt M, Youngstrom R, Hebert J, Hammond S, et al. A model for worksite cancer prevention: integration of health protection and health promotion in the WellWorks Project. Am J Health Promot. 1995;10(1):55–62.
15. Dow W, Philipson T, Sala-i-Martin X. Longevity complementarities under competing risks. Am Econ Rev. 1999;89(5):1358–71.
16. Montano D, Kasprzyk D. Theory of reasoned action, theory of planned behavior, and the integrated behavioral model. In: Glanz K, Rimer B, Viswanath K, editors. Health behavior and health education theory, research, and practice. 4th ed. San Francisco, CA: John Wiley & Sons; 2008. pp. 67–96.
17. Fishbein M. A reasoned action approach to health promotion. Med Decis Making. 2008;28(6):834–44.
18. Webb T, Sheeran P. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychol Bull. 2008;132(2):249–68.
19. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. Science. 1974;185(4157): 1124–31.
20. Gardner D. The science of fear. New York: Dutton; 2008.
21. Finucane M, Alhakami A, Slovic P, Johnson S. The affect heuristic in judgments of risks and benefits. J Behav Dec Making. 2000;13:1–17.
22. Witte K, Allen M. A meta-analysis of fear appeals: implications for effective public health campaigns. Health Educ Behav. 2000;27(5):591–615.
23. Janz NK, Becker MH. The health belief model: a decade later. Health Educ Q. 1987;11(1):1–47.
24. Lerner J, Keltner D. Beyond valence: toward a model of emotion-specific influences on judgment and choice. Cogn Emot. 2000;14(4):473–93.
25. Richardson A, Green M, Xiao H, Sokol N, Vallone D. Evidence for truth®: the young adult response to a youth-focused anti-smoking media campaign. Am J Prev Med. 2010;39(6):500–6.
26. Kahneman D. Thinking fast and slow. New York: Farrar, Straus & Giroux; 2011.
27. Loewenstein G, Weber E, Hsee C, Welch N. Risk as feelings. Psych Bull. 2001;127(2):267–86.
28. Loewenstein G. Out of control: visceral influences on behavior. In: Camerer C, Loewenstein G, Rabin M, editors. Advances in behavioral economics. Princeton, NJ: Princeton University Press; 2004. p. 717.

29. Ariely D, Loewenstein G, Prelec D. "Coherent arbitrariness": Stable demands curves without stable preferences. Q J Econ. 2003;118(1):73–106.

30. Wansink B. Mindless eating. New York: Bantam Books; 2006.

31. Wansink B, Park S. At the movies: how external cues and perceived taste impact consumption volume. Food Qual Prefer. 2001;12(1):69–74.

32. Pedersen S, Kang J, Kline G. Portion control plate for weight loss in obese patients with type 2 diabetes mellitus: a controlled clinical trial. Arch Intern Med. 2007;167(12):1277–83.

33. Wansink B, Van Itterum K. Portionsize me: Downsizing our consumption norms. Am Dietetic Assoc. 2007;107(7):1103–6.

34. Downs J, Loewenstein G, Wisdom J. Strategies for promoting healthier food choices. Am Econ Review. 2009;99(2):159–64.

35. Schwartz J, Riis J, Elbel B, Ariely D. Inviting consumers to downsize fast-food portions significantly reduces calorie consumption. Health Aff (Millwood). 2012;31(2):399–407.

36. Underhill P. Why we buy: the science of shopping. New York: Simon & Schuster; 1999.

37. Fuchs V. Time preference and health. In: Fuchs V, editor. Economic aspects of health. Chicago: University of Chicago Press; 1982. p. 93–120.

38. McClure S, Ericson K, Laibson D, Loewenstein G, Cohen J. Time discounting for primary rewards. J Neurosci. 2007;27(21):5796–804.

38a. Mischel W, Ebbesen EB, Zeiss AR. Cognitive and attentional mechanisms in delay of gratification. J Personality and Social Psychology 1972;21(2):204–18.

39. Gruber J, Koszegi B. Tax incidence when the individuals are time-inconsistent: the case of cigarette excise taxes. J Public Econ. 2004;88:1959–87.

40. Cherukupalli R. A behavioral economics perspective on tobacco taxation. Am J Public Health. 2010;100:609–65.

41. Gruber J, Mullainathan S. Do cigarette taxes make smokers happier. Adv Econ Anal Pol. 2005;5(1):1–45.

42. Gunby P. Canada reduces cigarette tax to fight smuggling. JAMA. 1994;271(5):647.

43. Ayres I. Carrots and sticks. New York: Bantam Books; 2010.

44. Giné X, Karlan D, Zinman J. Put your money where your butt is: a commitment contract for smoking cessation. Am Econ J Appl Econ. 2010;2:213–35.

45. Volpp K, John L, Troxel A, Norton L, Fassbender J, Loewenstein G. Financial incentive-based approaches for weight loss: a randomized trial. JAMA. 2008;300(22):2631–7.

46. Volpp K, Troxel A, Pauly M, Glick H, Puig A, Asch D, et al. A randomized, controlled trial of financial incentives for smoking cessation. N Engl J Med. 2009;360(7):699–709.

47. Cahill K, Perera R. Competitions and incentives for smoking cessation. Cochrane Database Syst Rev. 2011(4):CD004307.

48. Prochaska J, Redding C, Evers K. The transtheoretical model and stages of change. In: Glanz K, Rimer B, Viswanath K, editors. Health behavior and health education theory, research, and practice. 4th ed. San Francisco, CA: John Wiley & Sons; 2008. pp. 97–121.

49. Kim A, Kamyab K, Zhu J, Volpp K. Why are financial incentives not effective at influencing some smokers to quit? Results of a process evaluation of a worksite trial assessing the efficacy of financial incentives for smoking cessation. J Occup Environ Med. 2011;53(1):62–7.

50. Volpp K, Pauly M, Loewenstein G, Bangsberg D. P4P4P: an agenda for research on pay-for-performance for patients. Health Aff (Millwood). 2009;28(1):206–14.

51. Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. Econometrica. 1979;47(2): 263–91.

52. Chen M, Lakshminarayanan V, Santos L. How basic are behavioral biases? Evidence from capuchin monkey trading behavior. J Polit Econ. 2006; 114:517–37.
53. De Martino B, Kumaran D, Seymour B, Dolan R. Frames, biases, and rational decision-making in the human brain. Science. 2006;313(5787):684–7.
54. Banks S, Salovey P, Greener S, Rothman A, Moyer A, Beauvais J, et al. The effects of message framing on mammography utilization. Health Psychol. 1995;14(2):178–84.
55. Detweiler J, Bedell B, Salovey P, Pronin E, Rothman A. Message framing and sunscreen use: gain-framed messages motivate beach-goers. Health Psychol. 1999;18(2):189–96.
56. O'Keefe D, Jensen J. The relative persuasiveness of gain-framed and loss-framed messages for encouraging disease prevention behaviors: a meta-analytic review. J Health Commun. 2009;12(7):296–316.
57. Rothman A, Salovey P. Shaping perceptions to motivate healthy behavior: the role of message framing. Psychol Bull. 1997;121(1):3–19.
58. Meyerowitz B, Chaiken S. The effect of message framing on breast self-examination attitudes, intentions, and behavior. J Pers Soc Psychol. 1991;52(3):500–10.
59. Ariely D. Predictably irrational. New York: Harper Collins; 2008.
60. Ariely D, Loewenstein G, Prelec D. Tom Sawyer and the construction of value: Federal Reserve Bank of Boston Working Paper; 2005. Available from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=774970.
61. Finkelstein S, Fishbach A. When healthy food makes you hungry. J Consum Res. 2010;37:357–67.
62. Horgen K, Brownell K. Comparison of price change and health message interventions in promoting healthy food choices. Health Psychol. 2001;21(5):505–12.
63. Thaler R, Sunstein C. Libertarian paternalism. Am Econ Rev. 2003;93(2):175–9.
64. Camerer C, Issacharoff S, Loewenstein G, O'Donoghue T, Rabin M. Regulation for conservatives: behavioral economics and the case for "asymmetric paternalism." U Penn Law Rev. 2003; 151:1211–54.
65. Rayner G, Lang T. Is nudge an effective public health strategy to tackle obesity? No. BMJ. 2011;342: d2177.
66. Bonell C, McKee M, Fletcher A, Wilkinson P, Haines A. One nudge forward, two steps back. BMJ. 2011;342:d401.
67. Bonell C, McKee M, Fletcher A, Haines A, Wilkinson P. Nudge smudge: UK government misrepresents "nudge." Lancet. 2011;377(9784):2158–9.
68. Marteau T, Ogilvie D, Roland M, Suhrcke M, Kelly M. Judging nudging: can nudging improve population health? BMJ. 2011;342:d228.
69. Loewenstein G, Ubel P. Economics behaving badly. New York Times. 2010 July 15, 2010.
70. Loewensetein G, Asch D, Volpp K. Behavioral economics holds potential to deliver better results for patients, insurers, and employers. Health Aff (Millwood). 2013;32(7):1244–50.
71. Winickoff J, Gottlieb M, Mello M. Regulation of smoking in public housing. N Engl J Med. 2010;362(24):2319–25.
72. Volpp K. Paying people to lose weight and stop smoking. LDI Issue Brief. 2009;14(3):1–4.
73. Richards M, Sindelar J. Rewarding health food choices in SNAP: behavioral economic applications. Milbank Q. 2013;91(2):395–412.
74. Gneezy U, List J. The why axis: hidden motives and the undiscovered economics of everyday life. New York: Public Affairs Books; 2013.
75. Taylor N, Buttenheim A. Improving utilization of and retention in PMTCT services: can behavioral economics help? BMC Health Serv Res. 2013;13(1):406.
76. Zimmerman F. Habit, custom, and power: a multi-level theory of population health. Soc Sci Med. 2013;80:47–56.

CHAPTER 14

BIOLOGICAL PATHWAYS LINKING SOCIAL CONDITIONS AND HEALTH

Plausible Mechanisms and Emerging Puzzles

Laura D. Kubzansky, Teresa E. Seeman, and M. Maria Glymour

INTRODUCTION

This chapter considers mechanisms that may explain how social exposures "outside the body" get under the skin to influence physical health and disease. We conceptualize the many intersecting pathways by which social adversity can influence health as three broad categories: toxic environments (social or physical); health-relevant behaviors; and psychosocial stress and related cognitive/affective processes. These three mechanisms trigger a succession of biologic processes potentially related to health, discussed in more detail below (Figure 14.1). These categories are useful heuristics, but the pathways are not truly distinct and in fact are partly dependent on one another. The relationship between environments, behaviors, and cognitive/affective processes are characterized by feedback loops. The environment gives rise to external circumstances or events characterized as demands or stressors. These stressors may cause psychological and/or physical stress, which in turn leads to behavioral or physiological changes. According to this formulation, stress is experienced when individuals perceive that external demands exceed their ability to cope (for more detailed discussion of these relationships see Chapter 9). Behaviors and cognitive/affective processes of course also reciprocally influence environments, and nonsocial aspects of the environment also matter. However, in this chapter we are mainly concerned with describing the downstream biological processes that are likely mediated by stress-related processes; we do not additionally review the pathways linked to toxic physical exposures (e.g., chemical exposures at work, air pollution). The physiologic pathways potentially linking social

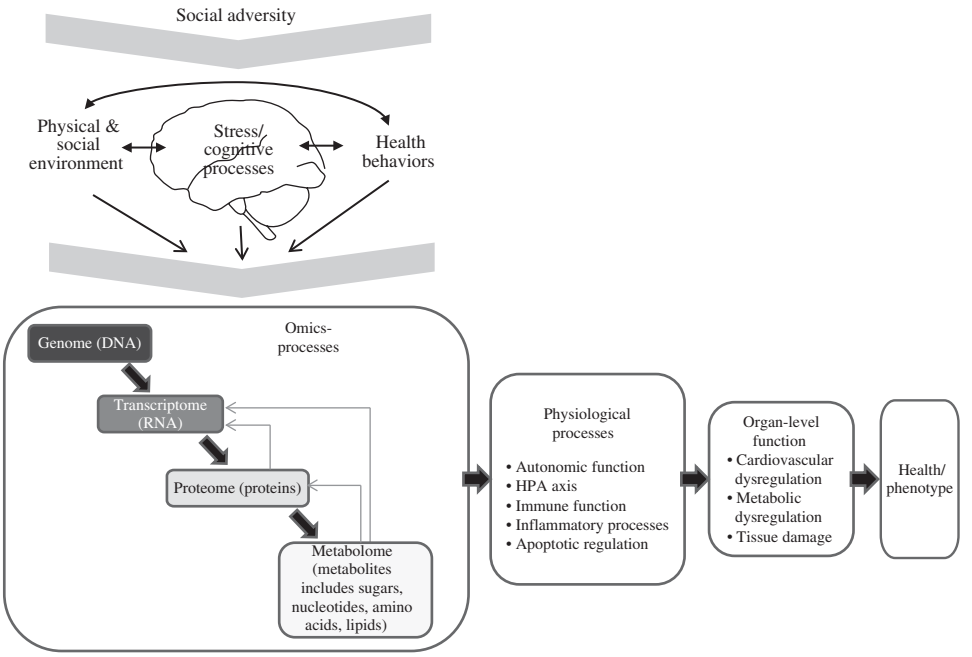


FIGURE 14.1: Biological pathways linking social conditions and health.

adversity to health are also complicated, intersecting, and interacting. The body responds to its environment as a living organism so that changes in one system trigger disruptions in other systems, which either singly or in combination may initiate pathophysiological processes. Moreover, the long-term health effects of dysregulation in any single pathway may depend on resilience or vulnerability that has already been established in other physiologic processes, presenting challenges for understanding biological effects of social exposures.

Despite this complexity, recent decades have brought tremendous progress in identifying the physiologic mechanisms linking social conditions to health and disease across the lifecourse. In this chapter, we consider biological processes posited to be influenced by "social adversity" broadly defined, and we use this umbrella term without emphasizing possible distinctions between specific physiological processes that might be differentially triggered by types of adversity such as poverty, social isolation, or discrimination. Although more detailed evidence may emerge in coming years, at this point the evidence base is largely insufficient to distinguish underlying pathophysiologic processes of one type of social adversity from another.

THE IMPORTANCE OF CONSIDERING BIOLOGICAL PROCESSES

Although social epidemiologic research is most concerned with identifying upstream determinants of population health, identifying the biological mechanisms linking social exposures to health is critical for guiding interventions, establishing causality, and motivating action.

Even when social exposures are identified as potentially toxic, we may not know how to design interventions for the greatest effectiveness. Uncertainty about interventions may revolve around both the *content* and *timing* of the proposed intervention. For example, a great deal of research suggests that depression predicts increased risk of a secondary event after an initial myocardial infarction. However, few psychosocial interventions targeting depression successfully reduced secondary event rates (1). One reason for the apparent discrepancy between observational and intervention studies may be that the interventions did not target the relevant etiologic window, due to inadequate understanding of the biological mechanisms linking depression and cardiovascular risk. Most interventions have been conducted among older adults, but there is little evidence on what duration of depression exposure is necessary to alter biological risk or how enduring the biological alterations may be (2). Thus the most powerful etiologic window for intervention is uncertain. Identifying biological mechanisms may help to identify a broader set of intervention opportunities and guide more effective and targeted interventions to break the links between social adversity and poor health.

Documenting how biological processes are dysregulated by adverse social exposures also provides convincing evidence for causation. As discussed in Chapter 2 and elsewhere in this text, there is ongoing debate as to whether social adversity and chronic stress are truly causal factors leading to poor health. Much of the evidence for the relationship is based on observational evidence, and a major concern is that we may have the causal direction wrong. Skeptics have argued that psychosocial stress and distress are byproducts of social adversity but do not, *in and of themselves*, explain how or why social conditions influence health. Furthermore, it is likely that underlying disease may increase an individual's likelihood of experiencing poor social conditions and stress, or another underlying factor (i.e., a genetic predisposition) may increase likelihood of experiencing both social adversity and disease states (3). This concern is corroborated by evidence that *acute* health events are costly and predict declines in socioeconomic status (4–6). Molecular- and cellular-level biological processes that do not substantially impair day-to-day functioning are less likely to influence socioeconomic disadvantage. Finding that social adversity predicts dysregulation in these biological indicators can therefore help establish that reverse causation—for example, poor health causing socioeconomic disadvantage—is unlikely to provide a complete explanation for social inequalities in health.

A final motivation for studying physiologic mechanisms is that biologically grounded explanations are often especially compelling to consumers of health research, such as policymakers. Descriptions of observed associations between social adversity and disease are often more powerful if accompanied by evidence for the biological pathways mediating the links (7, 8).

For social epidemiologists, it is important to keep in clear sight what may (or may not) be gained from research on the interplay between biological, psychological, and social processes. Numerous disciplines contribute to research linking biological processes to stress exposure and adversity, and only some of this research directly serves the goals of social epidemiology (9). Relevant evidence clearly indicates that numerous damaging biological alterations co-occur with psychosocial stress (10, 11) and that social adversity leads to such stress. Not all of the research to date has clearly contextualized the study of biological processes in terms of how social adversity impacts health. However, such research is critical for building a research base that can inform practice and policy. Our research should first demonstrate the direct biological mechanisms by which social disadvantage and chronic stress alter health-relevant

physiological processes and then use this knowledge to inform the development of effective interventions.

CASCADING EFFECTS OF PSYCHOSOCIAL STRESS AND BEHAVIOR

Evidence for health consequences of many behaviors and material conditions is nearly incontrovertible. Differential exposure to environmental factors known to be toxic (e.g., lead, air quality) are well documented (12, 13), as are differences in health behaviors (14, 15); moreover such known risk factors often appear to be especially damaging to individuals with other sources of social disadvantage (16). Although controversy remains, studies to date suggest that conditions such as reduced access to healthcare and differences in health behaviors (e.g., smoking) cannot fully explain health effects of social adversity (17). For example, in a study of British civil servants, cardiovascular disease risk increased as employment grade (a marker of SES) decreased, even though all individuals in the study had access to healthcare (18–21). As a result, without minimizing the significance of these other factors, the hypothesis that chronic psychosocial stress is another important pathway linking social adversity with health has been (and continues to be) strongly championed. Historically, investigators have distinguished between "direct" effects of psychosocial stress on physiologic dysregulation and indirect pathways mediated by behaviors. However, there is growing evidence that these pathways may not always be distinct.

Early research on the links between psychological stress and health identified a reliable biological stress response characterized by activation of the hypothalamic-pituitary-adrenal (HPA) axis and the sympathetic nervous system (SNS) (for more detailed description see Chapter 9; 22, 23). This initial neuroendocrine response is hypothesized to alter an array of downstream physiological parameters that may ultimately impair health. Scientists have identified biological effects of many types of stressors: acute time-limited stressors produced in laboratory settings; brief naturalistic stressors such as taking an exam; stressful single events with consequences that may subside over time, for example due to natural disasters; chronic stressors that require restructuring social roles, such as becoming a caregiver; traumatic experiences that can have long-term psychological effects; or simply frequent exposure to a varied set of stressors (24, 25). All of these types of stressors may result from social adversity and also lead to psychological stress. Regardless of stressor type, when individuals perceive external events (stressors) as overwhelming their capacity to cope, a sense of stress and negative affect results, which in turn triggers a biological stress response. Thus, stemming from observations of an acute biological response to stress, psychological responses to social adversity are hypothesized to trigger a cascade of pathophysiological processes that, when they occur repeatedly or without opportunity for restoration, may ultimately initiate disease-related processes (26, 27).

These psychological responses to stress can also trigger unhealthy behaviors (e.g., individuals who feel anxious may be more likely to smoke cigarettes or eat unhealthy diets) and may influence health via additional mechanisms. In fact, among those who accept the premise that psychosocial stress is causally related to physical health, there remains ongoing debate about whether effects are solely due to behavioral pathways or may also occur via direct biological alterations that in turn alter health outcomes (28). Further complicating the issue is recent work on biological

embedding, which has suggested that childhood exposure to adversity may lead to alterations in brain architecture that in turn change behavioral tendencies (29). Specific changes involve corticolimbic circuits that process stress and corticostriatal pathways that support self-regulation, which together lead to greater vigilance for threat, mistrust of others, and poor self-regulation of appetitive behaviors. These behaviors in turn are hypothesized to exacerbate proinflammatory processes implicated in a range of disease outcomes (e.g., diabetes, cardiovascular disease). Proinflammatory processes have been most widely examined to date, but future work may elucidate a broader array of health-relevant physiological processes likely impacted through these same alterations in brain architecture and function. Thus while they are often discussed as completely separate processes, behavioral and biological responses to stress are deeply intertwined.

CONCEPTUAL MODELS INVOKED TO EXPLAIN EMPIRICAL OBSERVATIONS

Several conceptual models have been proposed to explain the observed associations between social conditions and health or biomarkers of health. Closely related concepts have sometimes been developed in different research disciplines and thus have adopted inconsistent terminology. We have grouped these different traditions into two general categories: those that emphasize the accumulation of physiologic damage over time and those that focus on developmental timing of when exposures occur, positing a special importance of exposures during a particular (typically early) life stage. Most conceptualizations assume that individuals facing adversity—whether psychological or biological—make trade-offs and adaptations that may be beneficial in the short term but have undesirable consequences over the long run or in other contexts. However some models consider the ill health effects of adversity to be cumulative, building over time but without explicit reference to developmental timing issues, while other models invoke developmental evidence to suggest that effects may be particularly potent at specific developmental stages. In models that focus more on timing, poor health may be defined not only by the occurrence of disease or disabilities, but also by nonnormative timing of normative events (e.g., menarche, menopause).

HOMEOSTASIS, ALLOSTASIS, ROBUSTNESS, AND WEATHERING

To maintain internal integrity in the face of the regular and irregular challenges that result from interacting with the environment, organisms must react with appropriate biological changes. A systems-level approach to understanding how organisms manage these numerous biological changes has provided one framework for research on how and why social factors influence health (24). Initial work built on the concept of homeostasis, defined as a coordinated physiological process designed to regulate (or by some definitions, to resist) perturbations in the service of maintaining constancy or a single equilibrium point in key biological systems (e.g., blood glucose levels) (30). Homeostasis was conceptualized as the tendency to maintain the state of the system rather than its functions (31). Early work hypothesized that psychological or physical stress threatened homeostasis by altering the body's ability to maintain constancy.

However, investigators interested in the relationship between psychosocial stress and health became increasingly concerned that the notion of homeostasis did not sufficiently account for the continual biological changes and adaptations required to meet environmental challenges, even in the normal course of events. The concept could not easily accommodate the possibility that these changes might lead to a new steady state or take a toll on the system in some way. Perspectives to incorporate this possibility—including "robustness" and "allostasis" —were subsequently considered (31, 32). Robustness was initially proposed to describe fundamental and structural principles that govern and organize adaptive biological systems (31). Robustness is defined as a systems-level property that allows an organism to maintain effective functioning in the face of internal and external perturbations to the system (24). Unlike homeostasis, robustness is conceptualized as maintaining the functions of a system rather than maintaining a particular equilibrium point, and includes active consideration of feedback loops, redundancy, and diversity of function (31). Thus, robustness includes the capacity of the system to change its mode of operation in flexible ways in order to meet varying demands on the system. A system is considered robust as long as it maintains functionality, even if it must transition to a new steady state to do so, or if instability actually helps the system to cope with perturbations. When biological systems are repeatedly challenged, robustness may deteriorate, resulting in disequilibrium. Decline in robustness, or the capacity to equilibrate in the face of dynamic environmental challenges, is sometimes considered central to the loss of biological integrity associated with aging (31).

The notion of allostasis was developed along similar lines. Allostasis is defined as the dynamic processes through which biological regulatory systems alter their functioning in order to adapt to changing conditions/demands. Here the "stability" that is sought is that which allows the body to continue to function optimally under varying conditions. Such stability is achieved through change rather than through resisting change; in response to challenge, operating ranges of key biological processes may be altered (32). A key principle underlying the allostasis concept is the importance of efficiency. Organisms are designed to be efficient and trade-offs between systems may be required to achieve efficiency. Allostasis explicitly invokes the notion of adaptation and focuses on requirements for adaptation to current conditions, positing the need for system trade-offs to accommodate immediate needs. Homeostasis emphasizes the regulation of physiology within a fairly narrow and constant range (e.g., body temperature) focusing more on the end state of the organism; in contrast, allostasis focuses more on the process of responding to challenge by changing underlying physiological parameters in the short-term to maintain balance across systems (e.g., raising blood pressure or breaking down glycogen stores to release glucose into the blood stream). These frameworks provide the foundation for the two multisystem approaches that have dominated the field of research on stress and health to date.

Because theories of stress commonly posit that stress is a process by which biological systems are repeatedly challenged, exploring effects of exposure to stress has focused attention on the ability of the body to respond to environmental challenges and potential trade-offs between short-term adaptation and long-term costs (33). A systems-level approach further suggests the importance of considering multiple biological systems simultaneously rather than narrowly focusing in on a single system. Building on the earlier frameworks, two concepts that emphasize accumulation of adverse physical changes across multiple systems over the lifecourse have been proposed: allostatic load and weathering or accelerated aging.

Allostatic load refers to the "wear and tear" that accumulates in body systems when individuals are exposed to chronic stress. Wear on the system occurs when recurring stress responses activate the autonomic nervous system and neuroendocrine system to mobilize energy needed to adapt to immediate situational demands. These responses may be advantageous in the short run, but lead in the long term to biological damage (11). A myriad of evidence indicates that social disadvantage is associated with increased biological "wear and tear" (34). Moreover, social disadvantage has been linked with altered brain development and function, and these alterations can also drive downstream processes that increase allostatic load (22, 35). For example, corticolimbic pathways, which are composed of interconnecting structures such as the prefrontal cortex, hippocampus, and amygdala, are affected. These pathways regulate the HPA axis and SNS via the hypothalamus and adrenal gland and thus influence peripheral biology (34). Dysregulation in HPA and SNS activity correlate with alterations in cardiovascular, metabolic, and inflammatory parameters that have been linked with increased likelihood of developing a range of diseases.

The "weathering" literature similarly invokes the accumulation of physical and psychological adversity to explain the faster age-related deterioration of health in racial/ethnic minorities and low-SES individuals compared with the majority/advantaged groups. Exposure to higher psychosocial stress levels and relative lack of resources is hypothesized to manifest in physical deterioration at earlier ages as well as lead to the observed higher rates of chronic illness and lowered life expectancy in certain groups (36). The weathering literature proposes that under adverse circumstances, the optimal timing for certain developmental milestones or biological events occurs earlier. For example, Geronimus postulated that the nadir of infant mortality for black mothers occurs at earlier ages (10–20 years younger) than for white women and hypothesized that the high teen pregnancy rate among black women is an adaptive response to weathering.

Both allostatic load and weathering perspectives are systems-oriented in that they seek to capture functioning across multiple systems (often measured using multiple biomarkers) rather than considering only a single marker or endpoint. Most empirical work related to these perspectives has relied on assessing multiple clinical biomarkers and summarizing information across these components to characterize systemwide dysregulation.

BIOLOGICAL EMBEDDING/DEVELOPMENTAL ORIGINS OF HEALTH AND DISEASE

Biological embedding has been defined as occurring when social environments alter biological and developmental processes in enduring ways that influence the long-term health of the organism (37). This perspective incorporates a lifecourse orientation with the notion that early experience and its effects on development strongly influence later lifecourse health outcomes. Effects may occur regardless of intervening experience (e.g., latent effects), due to early experiences setting up a chain of events that influence later development and health, or due to cumulative effects of repeated exposures. Various biological mediators of "embedding" have been proposed, including epigenetic processes, tissue remodeling, alterations in neural structure and function, HPA axis effects, and shifting immune patterns toward inflammatory processes (38).

The concept of biological embedding was initially proposed by population health scientists to explain several observed epidemiologic patterns. In recent years, biological embedding has

become an organizing framework for much research on the interplay between social and biological factors, especially work focused on children. In addition to explaining the socioeconomic gradient in health (37), biological embedding theories aim to explain the following patterns: *(1)* while social inequalities in health are often evident early in life, they persist into adulthood and manifest in different biological systems as individuals age; *(2)* social inequalities in health are not fully explained by measured behavioral risk factors; *(3)* social inequalities differentially influence long-term health outcomes depending on when during the lifecourse adversity is experienced; *(4)* social inequality often appears to be intergenerational, with effects "transmitted" from parents to children.

Recent research has identified lasting effects of childhood social adversity on later life health and invoked the concept of biological embedding to explain these findings, specifically focusing on how psychosocial stress can program the response tendencies of cells involved in initiating and maintaining inflammation (29)—though it is also highly likely to impact other physiological processes as well. Numerous studies that seek to understand the interplay between biological and social processes implicitly invoke the biological embedding framework, although the developmental focus is often deemphasized, particularly in studies of adults. In such studies, the notion of embedding refers more simply to the idea that adverse experience is firmly fixed in the biology of the organism (38).

Imprinting models (unlike weathering or allostatic load) similarly emphasize a potential sensitive period during which adverse exposures are especially influential. In classical models of imprinting, the sensitive period occurs very early in life, although various physiologic mechanisms could theoretically create sensitive periods at other developmental stages. One model in this arena is the developmental origins of adult health and disease (DOHAD) model. DOHAD builds on the observation that the rapidly changing incidence of disease cannot be explained in terms of genetic alterations (39). The model suggests events that affect fetal growth can permanently alter the structure and physiology of the offspring in ways that increase risk of chronic disease in later life (40). Much of the emphasis in DOHAD literature has been on nutritional or other forms of material deprivation that affect fetal health. For example, DOHAD models suggest in utero nutritional deprivation leads to "fetal programming," whereby specific developmental paths are triggered that improve survival under calorie-poor conditions but have enduring consequences for cardiovascular risk. The very phenotypic accommodations that improve survival chances under calorie-poor conditions may increase cardiovascular risk under calorie-rich conditions. The fetal programming and DOHAD frameworks have been heavily influenced by research on the links between birthweight and adult health conditions (41) and findings of long-term health outcomes for cohorts conceived in famine conditions (42). Because social conditions shape nutritional and material deprivations, DOHAD has been invoked to help explain enduring social inequalities in health.

Fetal programming models emphasize that the physiologic adaptations made in response to in utero environments would be beneficial if the child's postnatal environment matched the prenatal environment. The differential susceptibility to context (DSC) hypothesis adds further complexity to the notion of biological imprinting occurring during child development (42, 43). According to this hypothesis, stress response phenotypes serve to calibrate a child's behavior to his or her likely environment, and the same phenotype is not most advantageous in all circumstances. According to a Swedish idiomatic expression, some children are "dandelions" (*maskrosbarn*), that is, they

have a low-reactive phenotype and a capacity to flourish within a large range of environmental circumstances, much like the eponymous weed that can thrive under varied soil conditions. By contrast, the "orchid" child (*orkidebarn*) has a high-reactive phenotype whose adaptation is highly context-dependent. The orchid is a flower of surpassing beauty in good conditions, but orchids wither under neglectful conditions; "orchid" children can be highly successful in positive social environments but fare very poorly in more adverse environments. Emerging evidence in this area is beginning to overturn the conventional wisdom that high stress reactivity is universally detrimental to health; rather it appears the effects of high reactivity on behavior and health may be bivalent, exerting either health-promoting or risk-augmenting tendencies depending on the context (43).

Empirical evidence is not yet available to test whether differential susceptibility to context alters health outcome across the lifecourse. However, the hypothesis has been invoked to suggest that children who are extremely responsive to context will have poor adult health if they experience a highly stressful early environment, but will have optimal adult health outcomes if they experience a highly supportive early environment. The hypothesis also suggests that under conditions of early social adversity, children who are less responsive to context may not inevitably have poor health in adulthood, but under highly supportive early conditions, also may not demonstrate as optimal health outcomes relative to children who are more sensitive to context. Thus, a model of DSC proposes a *qualitative* interaction between the environment and the child's characteristics and is distinct from the more commonly proposed diathesis-stress (or "dual risk") interaction models. A diathesis-stress model posits that the same contexts are optimal for all children, but children are differentially harmed by adverse contexts. Vulnerable children have some other background disadvantage (perhaps genetic), and when they encounter the adverse environment they cannot maintain good functioning. Diathesis-stress models do not acknowledge the possibility that the same phenotypic adaptations that make a child vulnerable in one context may be advantageous in another context. The DSC model emphasizes adaptive mechanisms related to neuroendocrine programming in infancy or early childhood, rather than in utero reorganization, which is the claim of "fetal programming" models.

In recounting the history of the concept of biological embedding Hertzman (37) notes three levels of biological embedding: experience and behavior, organ systems and cellular function, and gene function. One challenge for research is to incorporate insights obtained from the various levels and to synthesize knowledge across these domains. In later sections of this chapter, we summarize progress at each of these levels.

The key premise of biological embedding is that aspects of the social environment can alter biological function in predictable and enduring ways that have significant health consequences over the lifecourse (38). This conceptual framework underpins other approaches focused even more explicitly on understanding how health disparities develop and may be mitigated. For example, using what they term an "ecobiodevelopmental approach," Shonkoff and colleagues (44) provide a conceptual taxonomy of the types of stress response that may occur early in life, based on hypothesized differences in their potential to cause significant physiologic disruptions that may ultimately have health consequences. Using this categorization, a positive stress response is characterized as a dysregulated physiologic state that is

short-lived and mild to moderate in magnitude, and can be effectively managed with the help of a caring and responsive adult. A tolerable stress response is characterized by exposure to nonnormative experiences that are significantly threatening but that are buffered by supportive adult relationships. In contrast, a toxic stress response results from significant adversity and prolonged or strong activation of the body's stress response systems occurring in the absence of protective adult relationships. Such approaches more clearly emphasize the importance of cognitive and emotional processes (e.g., sense of control, capacity to regulate emotions) in the biologic embedding of social inequalities.

IMPLICATIONS OF ALTERNATIVE MODELS

Sensitive periods in early life, as hypothesized in "fetal programming" or biological embedding models, do not necessarily preclude the impact of cumulative processes such as allostatic load, but emphasize potentially disproportionate consequences of adverse exposures during particular periods of development. Some versions of biological embedding posit "critical" periods for certain systems, during which environmental conditions shape the developmental trajectory in irreparable ways; other models hypothesize that certain developmental periods are "sensitive" but some level of plasticity remains even after these developmental windows. This distinction is critical to guide interventions. If social disadvantage becomes biologically embedded during a critical period in early life, it suggests that resources for infants, children, and pregnant women may have large long-term returns, but investments in adults are unlikely to substantially reduce social inequalities in health. By contrast, weathering or allostatic load models would imply that each successive experience of social adversity causes accumulating physiologic damage, and interventions at any point in the lifecourse could be beneficial.

However, this issue may not be as simple as it initially appears. A key question for social epidemiology revolves around understanding whether biological alterations that occur as a function of social experience are irreversible or may be modifiable. There is a small but rapidly advancing field of study on critical periods. Research in neuroscience is beginning to suggest that there may be more plasticity throughout child and adult development than was previously thought. For example, neurogenesis is now well documented to occur in adult humans, contrary to long-standing teachings (45). Further, adult experiences have been shown to shape both recruitment of neural networks and structural features of the brain (46). Recently, work has also begun to identify molecular mechanisms that can contribute to reopening critical periods or enhancing plasticity. For example, one study of adult men considered ability to identify the pitch of a sound without a reference point, a skill that can only be acquired early in life. Findings suggested that men who were administered a histone-deacetylase inhibitor learned to identify pitch significantly better than those administered placebo, suggesting that the critical period neuroplasticity was regained (47). We do not yet fully understand how to trigger plasticity (and perhaps facilitate recovery from adverse exposures), or the extent of plasticity that can be triggered later in childhood and into adulthood; however, discoveries to date suggest that greater understanding of such processes may provide important clues to effective interventions for mitigating health disparities.

DEVELOPING MECHANISTIC UNDERSTANDING: PROGRESS IN THE "OMICS" ERA

Broadly speaking, research has progressed from focusing on more easily observed indicators of pathophysiology, such as clinical measures (e.g., heart rate, blood pressure) at the organ or tissue level, to more difficult-to-observe indicators that characterize biological mechanisms at the cellular or molecular level, for example, inflammatory markers, epigenetic markers, or gene expression. With the "omics" revolution and the rapid technological progress in capacity to measure a vast array of biological processes from genetics to proteomics and intricacies of transcription and translation to metabolomics, research on these processes is increasingly feasible.

In the realm of brain development, research has clearly demonstrated that development of the brain is an "experience-dependent" process. Experience activates specific pathways in the brain. Some existing connections are strengthened, new connections are made, and other connections may be attenuated based on use or lack of use of those connections (48). Similarly, research on biological embedding is based on the premise that biological profiles are also "experience-dependent." However, given the complexity (and still emerging understanding) of these biological processes, there are major measurement and methodologic challenges in linking cellular or molecular changes with higher level adaptational processes (e.g., psychological states, social interaction).

Studies of these relationships have taken a variety of forms, because alternative study designs have complementary strengths and limitations (see Table 14.1). Epidemiologic studies comprise a significant portion of the literature, whereby observational data are used to assess associations between social adversity or other exposures and indices of biological function. Much of the work to date has been cross-sectional due to difficulties in obtaining repeated measures of biological function. Laboratory or clinical studies have been conducted to evaluate acute physiological stress responses occurring in relation to some type of challenge. In addition, these relationships have also been explored using naturalistic studies that assess repeated measures of physiological function (e.g., blood pressure) in relation to ongoing day-to-day experiences (for comprehensive discussion of these methods see 49).

In biomarker based research, it is helpful to distinguish between the use of biomarkers as mere proxies of early disease risk and those thought to be mechanistically involved in the physiologic embedding of social factors. Many important biomarkers are established as proxies of a disease process, but are less clearly established as part of the causal pathway leading to the disease. For example, research has considered short-term cardiovascular reactivity to stress or reduced vagal tone as potential *mechanisms* driving increase risk of coronary heart disease (50–52). In contrast, many studies of emotional stress and C-reactive protein (CRP) acknowledge that CRP has not been demonstrated to be causally related to development of coronary heart disease but note that CRP nonetheless provides a useful marker of likely risk of developing cardiovascular disease (53). Thus, even if we show that social adversity causes elevations in CRP, it does not follow that intervening on CRP levels would help eliminate social inequalities in health.

From an empirical standpoint, the capacity to measure the genome has opened up a vast vista for expanding our understanding of the role of genetics in social inequalities. However, on the basis of current evidence—and despite tremendous investments in genome-wide association studies—only a tiny fraction of between-person differences in complex health conditions, health relevant behaviors, and psychological or social characteristics are explained by established genetic

TABLE 14.1: Alternative research designs

Research Design	Example	Strengths*	Weaknesses*
Randomized population trial	Romanian Orphan Study (183) Perry preschool experiment (266)	• Supports causal inferences, rules out reverse causation and confounding • Identifies a specific intervention	•Typically small samples, so wide confidence bands • Generalizability to other populations uncertain • Difficult to study whether variations on the original intervention would have similar effects
Laboratory-based experiments randomly assigning exposures in humans	Status manipulations (267) Randomly assigned to receive support during stressful situation (268)	• Supports causal inferences, rules out reverse causation and confounding • Establishes plausibility of physiologic mechanisms	• Limited range of exposures are ethical • Difficult to establish whether short-term responses translate into long term health consequences • Generalizability to other populations uncertain
Laboratory-based experiments in animals	Rat pup-handling experiments (269)	• Supports causal inferences, rules out reverse causation and confounding • Can sometimes directly manipulate host (e.g., knock-out mice) to test specific causal mechanisms • Establishes plausibility of physiologic mechanisms •Variations on the definitions of exposure, e.g., differences in timing or content of exposure can be evaluated	• Limited range of exposures are ethical • Difficult to identify appropriate animal models for range of human experiences and human physiology • Difficult to determine how accurately animal models translate to human experience and biology
Natural experiments	Social policy changes (270) Economic fluctuations (271)	• Somewhere between observational evidence & RCTs in being able to (somewhat) rule out reverse causation and confounding • Can sometimes be used to evaluate differences in timing • Social policies correspond with specific, feasible interventions • Can lend insight into biological mechanisms	• Limited range of exposures correspond with the natural experiment, since it was not under investigator control • Often wide confidence intervals because most people in the population were not affected by the natural experiment or were affected to only a modest extent • Depends on assumptions for the natural experiment, and, if analyzed incorrectly, loses the advantages of "pseudo"-randomization

(*continued*)

TABLE 14.1: Continued

Research Design	Example	Strengths*	Weaknesses*
Observational cohort studies of social adversity and disease risk	Whitehall study of British civil servants (20)	• Longitudinal design rules out the simplest version of reverse causation • Large, representative samples	• Vulnerable to bias if early health or risk factors affect social risk • Exposure may not clearly correspond with a feasible intervention
Observational studies of social adversity and biomarkers of disease risk	Biomarkers in ongoing cohorts, such as the Normative Aging Study (272), MacArthur studies of successful aging (273)	• Rules out subtler versions of reverse causation if biomarkers do not impair functioning • Can use large, representative samples • Can help establish biological mechanisms, if biomarker is believed to be causal	• Vulnerable to bias if early health or risk factors affect social risk • Exposure may not clearly correspond with a feasible intervention • Expense and burden of biomarker assessments often (but not always) entails smaller, nonrepresentative samples • Biomarkers often (but not always) only measured once, exacerbating causal inference challenges.
Observational studies of social adversity and proteome, metabolome, transcriptome, or genome	Microarray analysis of gene expression patterns in lonely compared to nonlonely individuals, and bioinformatics analyses of promoters (225)	• Can help establish biological mechanisms • Can reveal biological pathways and mechanisms	• In untargeted analyses (e.g., metabolomics or genome-wide association studies), vulnerable to chance findings and false positive results • Uncertainty about underlying physiology often results in ambiguous results

* Strengths and weaknesses are described with respect to understanding physiologic bases for social inequalities in health, with a goal of informing strategies for eliminating inequalities.

variants. There is an unexplained gap between estimates of heritability—that is, the similarity between people with similar genetic backgrounds—and the measured importance of any specific, identified genetic polymorphisms to the phenotype. This finding leads to two conclusions. First, it is unlikely that most associations between social conditions and health are largely attributable to confounding by genetic background and certainly not due to single high-impact genetic polymorphisms. We do not anticipate discovery of the "poverty" gene. Although we do not rule out a role for genetics, the relevant genetic effects are likely dispersed over a large number of alleles, each with small individual effects.

Second, current research provides suggestive evidence that genetic background has different effects depending on environmental context (i.e., gene-by-environment interaction). This is consistent with the Differential Susceptibility to Context and fetal programming theories (54). Adaptive phenotypic plasticity implies that the characteristics that make some individuals vulnerable to adversity ("the orchid child") may also help them flourish in advantageous environments; as a result, these individuals have increased sensitivity to both negative and positive influences.

GETTING INSIDE THE BODY: PHYSIOLOGIC EFFECTS OF SOCIAL ADVERSITY

In this section we highlight evidence on how the social environment alters biological functioning in health-relevant ways. Our focus is on mechanisms plausibly related to the biological stress response per se. We describe specific physiologic responses that could underlie the theoretical ideas described in the models of allostatic load, weathering, or biological embedding frameworks as described above. We draw on recent discoveries in basic biology (i.e., technological advances with the "omics" revolution) as well as research focused on biological measures at the organ and tissue level to consider current work integrating biological processes with psychosocial exposures. We begin with a brief review of findings with what we call clinical markers of pathophysiologic responses, including biological measures of cardiovascular, metabolic, and endocrine processes. Next, we consider research that assesses potential multisystem effects of these exposures. We then consider evidence for the role of more granular processes, including alterations in the transcriptome and genome. See Figure 14.1 for a schematic overview of how social adversity might get under the skin to influence biological processes from the cellular level through to organ level function and ultimately lead to manifest disease.

CLINICAL INDICATORS

The majority of social epidemiological research has considered the relation between social adversity or other social exposures (e.g., chronic stress) and alterations in a range of measures of biological function that serve as early markers of disease risk. These links have been documented across the lifecourse, from birth through old age. The accumulated evidence suggests a cascade of events, which may begin prior to conception with epigenetic patterns reflecting parental exposures such

as diet. Gestational and early childhood conditions including nutrition, infection, and psychosocial stress, can influence cell and tissue development in the cardiovascular, neurologic, and immune systems.

BIRTHWEIGHT/SMALL FOR GESTATIONAL AGE

A large literature documents social inequalities in risk of low birthweight and newborns who are small for gestational age (SGA) (55), with inequalities in prenatal conditions hypothesized to be precursors of adult health inequalities. Low birthweight and SGA have established consequences with respect to infant health, and according to DOHAD models of "fetal programming" they are likely involved in the evolution of lifecourse inequalities in health. Empirical research has linked adverse conditions in utero to adult health, particularly cardiometabolic disorders (56, 57). Early evidence of this link was drawn from ecological studies. In the mid-1980s, Barker and colleagues reported that in 212 regions of England and Wales, there was a correlation of 0.73 between infant mortality rate in 1921–1925 and ischemic heart disease mortality rate in 1968–1978 (58). They concluded that undernutrition during prenatal or early postnatal periods increased later risk of metabolic dysfunction (obesity, insulin resistance) as well as ischemic heart disease. This initial ecological work was recognized to have many limitations, but launched an avalanche of studies to further evaluate the hypothesis. Since the first publications on fetal programming, individual level longitudinal studies have replicated the major result; meta-analyses suggest each additional kilogram of birthweight is associated with a 12% reduction in cardiovascular mortality risk (59).

These findings contributed to development of the fetal programming theory that suggests in utero exposure to undernutrition leads to the programming of a "thrifty" phenotype, which increases the risk of obesity and insulin resistance later in life. Crucially, the consequences of this early programming appear to hinge on the "mismatch" between the fetal nutritional environment and the environment after birth. Thus, individuals exposed in utero to privation during the Dutch Famine Winter (at the end of the Second World War) were more likely to become obese and to develop insulin resistance as adults, presumably because of the mismatch between their fetal environment and the postwar nutritional environment in Holland. By contrast, individuals exposed in utero to famine during the Siege of Leningrad were no more likely to become obese in adulthood, presumably because hunger persisted into the Soviet era.

A major challenge in this literature is that the biological signature of the thrifty phenotype has yet to be documented, though it is assumed to be mediated by epigenetic mechanisms. It is additionally unclear that gestation per se is the critical period; some evidence suggests intergenerational transmission of disadvantage via the accumulated lifelong effects deprivation experienced by the mother, even prior to conception (56). In combination, nutritional deprivation is thought to lead to a profile of reduced cell growth in the kidneys and pancreas, decreased insulin sensitivity, greater lipid storage, and increased cardiovascular reactivity (60). Major questions remain regarding critical periods of development (e.g., preconception, first trimester), the exact nature of adverse uterine conditions that have enduring effects (e.g., caloric restriction or protein restriction), and the role of "catch-up" growth.

CARDIOMETABOLIC FUNCTION

As reviewed in previous chapters, various dimensions of social adversity have been linked with incidence of major clinical cardiovascular outcomes, including acute myocardial infarction, sudden cardiac death, stroke, and atrial fibrillation (61–66). These associations are pervasive across diverse contexts and using a range of indicators of social adversity. Childhood SES strongly predicts coronary heart disease and stroke (67), supporting the likely importance of early life developmental periods. Further, migration studies suggest that people born in high stroke-risk areas who later migrate to lower risk regions nonetheless have high stroke risk in adulthood (68, 69). Interestingly, this work is not clear about the most relevant timing, and some evidence suggests that later childhood or adolescence is a sensitive period (69). Although we do not distinguish here between types of social adversity, the finding of specificity between early life predictors and various cardiac outcomes suggests multiple distinct, possibly even offsetting, physiologic pathways culminating in complex outcomes such as stroke. For example, although low birthweight is robustly associated with coronary heart disease, high birthweight is associated with increased risk of atrial fibrillation, which is itself a leading risk factor for stroke and other cardiovascular outcomes (70).

Because evidence for a link between social adversity and stress has been strongest in relation to cardiometabolic diseases, biomarkers in this domain have received especially intensive attention. Thus, studies have frequently considered social (e.g., SES) and psychological (e.g., depression) exposures in relation to major cardiovascular risk factors, for example, resting blood pressure, glucose control or insulin levels, and lipids (including triglycerides and total, high-density lipoprotein [HDL], and low-density lipoprotein [LDL] cholesterol) (e.g., 18, 71, 72). Social inequalities in subclinical indices of cardiovascular disease including carotid intima media thickness (IMT), coronary artery calcification (CAC), arterial stiffness, and indices of the vascular endothelium (e.g., arterial stiffness as measured by aortic pulse wave velocity and flow mediated dilation) are also well documented (34). For example, one study of US African American and Caucasian adolescents found indices of low SES (e.g., lower parental education levels, fewer socioeconomic assets) were associated with greater IMT and arterial stiffness, with racial disparities also apparent (72, 73).

Additional evidence on the physiologic consequences of social stressors has emerged from experimental studies. For example, one study brought men and women in good health drawn from a cohort of British civil servants into a laboratory and monitored cardiovascular measures during performance of two behavioral tasks, and for 45 minutes following a stress induction protocol (74). Among individuals with a lower employment grade, the return of blood pressure and heart rate variability to resting levels was less complete after 45 minutes, relative to individuals with a higher grade of employment. The odds of failure to return to baseline by 45 minutes in the low relative to the high grade of employment groups were 2.60 (95% CI: 1.20–5.65) and 3.85 (1.48–10.0) for systolic and diastolic pressure, respectively, and 5.19 (1.88–18.6) for heart rate variability. Thus, the investigators suggested that lower SES is associated with delayed recovery in cardiovascular function after mental stress. The laboratory-based set-up of this study design is powerful but nonetheless observational in the sense that the primary exposure of interest (SES) was not experimentally manipulated.

Even more compelling laboratory-based studies take advantage of subtle psychological cues (e.g., based on how a study confederate interacts with the study subject) that are known to

alter an individual's sense of social status. Such tools allow us to accomplish in the laboratory what could never be done in community-based population studies: we can randomly assign social status. For example, Mendelson et al., randomly assigned 44 healthy women to either an induced-subordinate or an induced-dominant status condition. Compared with induced-dominant status, induced-subordinate status increased negative affect and systolic blood pressure over the course of the study. These results suggest that even a short-term induction of subordinate status can have adverse effects on stress-related physiological systems (75). The tremendous strength of such a study is that it eliminates concerns of possible confounding because the primary exposure of interest (perceived dominant or subordinate status) is randomly assigned. The limitation is that we do not know if the short-term changes observed in the laboratory persist and accumulate into long-term disease; for this, we must rely on other types of study designs.

Animal models corroborate the human evidence. For example, animal research finds substantial metabolic changes in relation to induced chronic stress, with alterations in amino acid, carbohydrate, and lipid metabolism. In animals exposed to chronic stress, glutamate (a molecule involved in synthesizing glucose) levels were significantly or suggestively increased and glutamine (also involved in regulating glucose) levels significantly decreased when measured in blood (76, 77), urine (78), brain (79), and heart (80). Such metabolic changes have also been associated with cardiovascular disease and diabetes in other animal studies. For example, glutamine supplementation in mice was associated with increased glucose tolerance and decreased blood pressure (81).

IMMUNE FUNCTION/INFLAMMATORY PROCESSES

Chronic elevations in inflammation processes predict significant health risks, including cardiovascular disease, diabetes, and cognitive decline (51, 82, 83); however, it is not yet established which inflammatory markers are causal mechanisms leading to a given disease (84). Although social predictors of many markers of immune function have been evaluated, CRP, a proinflammatory molecule that strongly predicts a host of cardiovascular outcomes (85), has received the most attention.[1]

Disturbed immune function and dysregulated inflammatory response have been robustly linked with social disadvantage and with chronic stress and adversity in adults (25), and resistance to infection by common cold virus is higher among individuals with good social integration and support (91). In addition, numerous studies have documented relationships between early adversity, such as childhood maltreatment (92) or family SES (93, 94) with immune

1 In addition to CRP, commonly evaluated markers include cytokines (e.g., interleukin-4 [IL-4], interleukin-6 [IL-6] interferon-c [IFNc], tumor necrosis factor-α [TNF-α]), fibrinogen, and natural killer (NK) cell function. CRP modulates leukocyte activities and endothelial functions, and increases inflammation by catalyzing monocyte activation (86, 87). NK cells are lymphocytes involved in cell-mediated immunity. NK cells are thought to promote atherosclerotic lesion development and to be involved in tumor immunosurveillance (87–89). Proinflammatory cytokines are regulatory peptides that promote systemic inflammation, and can affect the expression of adhesion molecules, endothelial permeability, lipid metabolism, and other processes relevant to the development and progression of CVD (87). Fibrinogen is a protein that contributes to clotting of blood and blood thickness, is integral to thrombolytic events that trigger stroke and myocardial infarction, and can indicate the level of inflammation occurring in the body (86, 90).

dysregulation. Such effects have been demonstrated with a variety of markers of inflammation including elevated levels of CRP and IL-6 and other cytokines, as well as with fibrinogen and TNF-α (18, 95). Much of this research, however, has considered childhood circumstances in relation to levels of inflammation as measured in adulthood, and most studies relied on retrospective reports of childhood circumstances. Although the majority of studies use inflammatory indicators at a single point in time, several studies have used repeated measures of psychological distress and inflammation. These results indicate that while distress predicts subsequent inflammation, inflammatory levels were not associated with increased likelihood of being distressed at subsequent follow-up assessments (96, 97). Further, meta-analysis of experimental studies, in which participants are exposed to acute psychological stressors in a laboratory setting, indicate consistent elevations in Il-6 and Il-1β, though responses in other measures were less consistent (98).

Recent studies have also evaluated relations of social adversity with immune and inflammatory processes assessed earlier in the lifecourse. Such work often seeks to address one of two key issues: how early in life dysregulation in immune processes become evident, and whether early dysregulation predicts subsequent disease (e.g., CVD) risk in adulthood. For example, one study of British children found exposure to adverse events prior to age 8 was associated with elevated inflammation (as assessed by CRP and IL-6 levels) at age 10 and in midadolescence, with associations partially mediated by body mass index (99). A systematic review of studies of early social adversity and CVD-related immune biomarkers in children and adolescents found that associations were somewhat inconsistent but more often observed in studies with large samples (100). As noted in this review, some work suggests that among young healthy individuals, exposure to adversity amplifies the inflammatory response to exogenous immunologic stimuli but does not alter basal levels of inflammation. However, over time, the consequences of repeated hyperinflammatory responses may accumulate and lead to chronically elevated levels of inflammation (100). Thus, studies in younger individuals may see stronger evidence of immune and inflammatory dysregulation when examining response to exogenous stimuli as compared with assessing circulating concentrations of inflammatory biomarkers. For example, several studies have reported that harsh family climate is significantly related to cytokine response to lipopolysaccharide exposure (an exogenous bacterial stimulus) but not with ongoing inflammatory activity (i.e., circulating levels of IL-6) under normal conditions (101, 102).

Other work has considered effects of psychosocial stress or various forms of social adversity on wound healing. Wound healing entails a complex sequence of events largely regulated by the cellular immune system; these processes seem to be responsive to psychological and social exposures (103). As a clinical marker, wound healing provides a relevant outcome in and of itself (e.g., recovery from injury or surgical procedures). Most of this work assesses the role of chronic stress on biological mechanisms associated with acute injury; this evaluates a different set of processes from those typically studied with health outcomes that develop over a longer period. Studies in this domain assess effects on ongoing chronic psychosocial stress (naturally occurring) in response to an induced wound (e.g., via punch biopsy). A relatively recent systematic review and meta-analysis reported that a majority of studies found chronic psychological stress was associated with impaired wound healing or dysregulation of a biomarker related to wound healing (104). For wound healing, increased expression of proinflammatory cytokines is initially protective. Impaired wound

healing in response to ongoing chronic psychosocial stress has been attributed to an attenuated inflammatory response in the initial phase of the repair process (105). One explanation for this has been stress-related overproduction of glucocorticoid hormones, which reduce expression of cell adhesion molecules involved in immune cell trafficking. Such research has helped to clarify that chronic stress can impair immediate immune response to injury by suppressing aspects of adaptive immune function (106).

HYPOTHALAMIC-PITUITARY-ADRENAL AXIS

The hypothalamus, pituitary gland, and adrenal gland (referred to as the HPA axis) respond in a coordinated fashion to environmental stimuli, communicating via endocrine hormones secreted into the blood stream. The HPA axis along with the sympathetic adrenal medullary (SAM) axis, are two principal systems mobilized in response to stress. As an individual interacts with the environment, stimuli encountered may serve as challenges or stressors that elicit responses from the HPA axis (as well as other internal homeostatic systems). HPA activation involves release of corticotrophin releasing hormone (CRH) from the hypothalamus to the anterior pituitary gland, which in turn releases adrenocorticotrophic hormone (ACTH) into the circulation. When ACTH reaches the adrenal cortex (located atop the kidneys), it stimulates the release of cortisol (107). Cortisol is a glucocorticoid, a steroid hormone that binds to the glucocorticoid receptor. Cortisol acts in a negative feedback loop by inhibiting further production of ACTH (107).

Laboratory studies of acute psychological stress suggest that cortisol response is more pronounced in situations that are highly uncontrollable or socially threatening (108). Chronic psychosocial stress and adversity has also been linked with dysregulated cortisol levels (109, 110). In addition to cortisol, the adrenal cortex under ACTH stimulation also releases dehydroepiandrosterone (DHEA), a precursor to the sex hormones testosterone and estrogen, as well as aldosterone. Aldosterone is a mineralocorticoid, a steroid hormone that binds to the mineralocorticoid receptor and maintains blood volume and blood pressure by controlling the sodium/potassium balance of the blood. Biologically, DHEA acts as a cortisol antagonist, buffering the actions of cortisol, and altering the sensitivity of glucocorticoid receptors (111). DHEA has neuroprotective properties; it can modulate the vulnerability of an organism to negative consequences of stress (112). Under nonstress conditions, DHEA and cortisol are correlated; however, during chronic stress, DHEA and cortisol patterns become dysregulated (111). Thus, adrenal output of DHEA and the cortisol:DHEA ratio can provide important information regarding the biologic consequences of stress (113). Dysregulated levels of cortisol have been linked with greater risk of a number of poor health outcomes including cancer, heart disease, and diabetes (114–116). These associations are hypothesized to be influenced by downstream effects of excess glucocorticoid exposure (resulting from chronic stimulation of the HPA axis) that include immunological and metabolic effects. For example, high levels of glucocorticoids are associated with changes in innate and adaptive immune responses as described above (113). This altered insulin signaling leads to increased likelihood of insulin resistance, hyperlipidemia, and visceral adiposity (117).

Cortisol has become one of the most commonly examined biomarkers of the stress response, possibly because of its ease of measurement (with noninvasive techniques such as sampling of saliva). However, the interpretation of cortisol measurements is complex because of its diurnal

fluctuations, lack of reference values, sensitivity to extraneous stimuli (e.g., smoking, eating, or just getting out of bed in the morning), and heterogeneity in the observed pattern of response to chronic stress (e.g., cortisol secretion can be either elevated or flattened, depending on the type of stress). The diurnal rhythm of cortisol peaks in the morning and has a nadir in the evening. Chronic stress tends to blunt diurnal cortisol rhythms, lowering levels in the morning but increasing them in the evening (107), with a net effect of increasing total daily cortisol output (107). For example, studies have demonstrated higher levels of cortisol associated with more difficult family environments (118) and lower SES (e.g., 119, 120). However, research has also indicated that the effects of psychosocial stress on cortisol are nuanced and results often appear inconsistent (121). The type of stressor can influence the magnitude of the increase in cortisol output (107). Further, there is a complex relationship between stress levels and cortisol whereby high distress levels are associated with increased cortisol output, but some forms of extreme distress (e.g., post-traumatic stress disorder) are associated with reduced output, suggesting a failure of the neuroendocrine system to respond appropriately to stress signals (109, 110).

DHEA also has a diurnal rhythm, and in animal models, higher levels of DHEA were associated with enhanced immune function and memory, reduced depression and anxiety, and lowered corticosterone-induced performance-related decrements (122, 123). Accurately capturing cortisol and DHEA diurnal rhythms requires sample collection at repeated intervals throughout the day, which is difficult to implement in large population-based studies. Recently, assays for measuring cortisol and DHEA have been developed in toenails and hair that reflect long-term cortisol and thereby perhaps may provide a better marker of chronic psychosocial stress (122–124).

HPA activation, through the renin-angiotensin pathway, also stimulates the release of aldosterone, which is in the same biologic pathway as cortisol and DHEA (125). Although aldosterone has not been as extensively studied as cortisol or DHEA in relation to psychosocial adversity or stress, there is evidence to suggest it might provide insight into the biological underpinnings linking these exposures with poor health. Several studies have suggested that aldosterone is higher in depressed versus nondepressed patients (126, 127), and lower in those who practice stress reduction techniques (128). Animal studies consistently demonstrate altered aldosterone levels in response to stress (125). Moreover, aldosterone is associated with increased oxidative stress and inflammation (125, 129). Although aldosterone has a modest diurnal rhythm, one blood measurement appears to reflect levels over time, as evidenced by its association with various chronic diseases (130, 131) in prospective studies. One reason that aldosterone may be less studied than cortisol is that past assays required large volumes and a difficult extraction step. This barrier may be overcome with the recent development of a high-throughput, low-volume liquid chromatography/mass spectrometry (LC-MS/MS), an assay that has been demonstrated to have excellent precision (132).

AUTONOMIC FUNCTION

The autonomic nervous system includes sympathetic and parasympathetic nervous systems (SNS and PNS), each a central regulatory system assisting in maintaining the body's overall physiological integrity in the face of varying environmental stimuli. The SNS is mobilized in response to challenge and triggers "fight or flight" responses. The PNS is responsible for activities that occur

when the body is at rest, which have been characterized as "rest and digest." In the context of stress response, activation of the PNS may promote faster recovery from challenge and mechanisms of restoration. Most studies that consider parasympathetic function in relation to social status have used measures of heart rate variability (HRV) to assess balance between the SNS and the PNS in regulating the heart. Reduced HRV is related to either *increased* sympathetic or *decreased* parasympathetic autonomic regulation. Lower HRV predicts ventricular arrhythmia and sudden death among patients with diagnosed CHD, independent of other predictors of sudden death (133, 134). Reduced HRV is also a risk factor for sudden cardiac death among individuals without diagnosed CHD (135, 136). Heart rate variability provides a noninvasive measure of cardiac autonomic tone and can be assessed by R-R intervals (a standard method of measuring heart rate based on measuring the time intervals (in milliseconds) between consecutive peaks of the R waves on the QRS complex as recorded by an electrocardiogram (137, 138).[2]

A number of studies have linked various forms of chronic stress with reduced HRV in both adults and children (139–141). For example, an inverse gradient of HRV has been demonstrated with work-related stress as well as with various forms of distress including anger, depression, and anxiety (142–145). Other work has suggested chronically impaired autonomic function is also associated with low social status. For example, in 2,197 middle-aged male civil servants, low employment grade was associated with low HRV (146). While further research has demonstrated associations between low SES as measured by other indicators (e.g., educational attainment) and low HRV (147), evidence also suggests that associations may differ by race/ethnicity (148). A clear gradient in HRV by level of psychological stress and has been established (149) but the evidence is less clear for social status as a predictor of HRV, in part because most studies are cross-sectional, making causal inference problematic (e.g., 147).

Activation of the SNS results in the release of the catecholamines, epinephrine and norepinephrine, from the adrenal medulla. In animal models, chronic stress leads to continuous activation of the SNS and increased circulating norepinephrine levels (150–152). In humans, studies have demonstrated that circulating norepinephrine levels are higher among individuals who have experienced childhood abuse as well as those with chronically high distress (e.g., 153, 154). Evidence on the relationship between catecholamine levels and SES is mixed, with graded associations for both epinephrine and norepinephrine found in the CARDIA (155) and smaller samples (156, 157), but other studies reporting no or inconsistent associations (158, 159). The limited work and mixed results may reflect the challenges of using catecholamines as proxies for physiologic stress. Basal levels of these hormones vary across individuals, with a tenfold biologic range (160). Although norepinephrine levels increase in blood within minutes of an acute stress, circulating blood levels in an individual at rest may not represent a cumulative measure of psychosocial stress. Although norepinephrine assessed in urine likely provides a more time-integrated

2 In addition to the time domain measures of HRV (based on calculation of R-R intervals), it is also possible to calculate frequency domain measures from mathematical manipulations performed on the same ECG-derived data. The approach uses Fourier transforms. The HRV spectrum contains two major components: the high-frequency (0.18–0.4 Hz) component, which is synchronous with respiration and is identical to respiratory sinus arrhythmia (RSA). The second is a low frequency (0.04 to 0.15 Hz) component that appears to be mediated by both the vagus and cardiac sympathetic nerves. The power of spectral components is the area below the relevant frequencies presented in absolute units (square milliseconds). Reduced LF power is believed to be of prognostic significance.

exposure (161), few large-scale epidemiologic studies collect urine. Thus, reliable evidence on social adversity and chronic alterations in catecholamines is limited.

CELL STRUCTURE AND TISSUE REMODELING

At the cellular level, social context is hypothesized to influence both cell division and cell death (e.g., via apoptosis), as well as structural features of the cell such as dendritic shape and synapse formation. In undifferentiated cells, environmental stimuli alter patterns of gene expression to set the cell's developmental trajectory on a specific path. Related processes occur during both development and regeneration or repair (162), and determine the composition and size of the tissues or systems composed of those cells. For example, the SNS innervates the bone marrow microenvironment where hematopoiesis and monocyte differentiation occur, and cell differentiation is moderated by norepinephrine signaling. Recent experiments indicate that SNS stimulation, triggered by social stress (in humans) or social defeat (in mice) increases monocyte differentiation, tipping the immune system toward a proinflammatory emphasis (163). Our understanding of the effects of the social environment during development is in its nascence, but physiologic changes induced by social stimuli may alter biological functioning by means of tissue development or remodeling. Such changes can modify the sensitivity of the remodeled tissue to subsequent environmental stimuli, and these alterations may persist over the life of a cell and beyond (106).

BRAIN STRUCTURE AND FUNCTION

Changes to either the structure or function of neural tissue and circuitry may be especially relevant for mediating health effects of social adversity, because the brain plays a central role in regulating behavioral and physiologic responses to the environment. Animal models have provided the strongest evidence to date of the biological plausibility of long-term changes to brain structure or function induced by environmental context, and evidence from humans is now accumulating (164). Major challenges in this literature relate to establishing the causal direction between adversity and neurologic measures, because compelling longitudinal assessments of neurologic measures are rarely available. For example, individuals with major depressive disorder have demonstrated reduced anterior caudate volume (165) and reduced nucleus accumbens responsiveness to rewarding stimuli (166), but it is not entirely clear whether volumetric differences are causes or consequences of depression. Similarly, in a study of middle-aged Australian adults, self-reported current financial hardship correlated with smaller hippocampal and amygdala volume, but retrospectively reported childhood poverty was not associated with volumetric measures (167). Many domains of SES also correlate with white matter tract integrity as measured with diffusion tensor imaging (168). In addition to measures of brain structure, social factors are also correlated with functioning of neural systems. For example, retrospectively reported lower parental social standing correlates with greater amygdala reactivity (169).

Prospective studies have helped address the challenge of establishing causal direction. For example, Gianaros et al., followed 48 healthy women from 1985 to 2004 and found that life stress reported during this period predicted hippocampal gray matter volume assessed in 2005 and 2006

(170). Natural experiments also bolster evidence that the causal direction is from stress to brain structure. In a small study of 17 adults "exposed" to the September 11 attacks (i.e., living within 1.5 miles of the World Trade Center on September 11, 2001) and 17 adults who were "unexposed" (living elsewhere at the time), Ganzel and colleagues assessed gray matter volumes several years after the World Trade Center events. They found that exposure predicted reduced gray matter volume in several brain regions, including the anterior hippocampus, medial prefrontal cortex, and amygdala (171).

McEwen hypothesizes a "glucocorticoid cascade" explanation for brain aging, based on the observation that stressful experiences trigger adrenal steroid release (as described above). Hippocampal neurons are thick with receptors to glucocorticoids, and repeated exposure to stressful events causes adverse remodeling of cell structure (change in dendritic shape and density). These effects, along with effects on the prefrontal cortex (172), may provide a neurobiological basis for evidence that stress causes transient cognitive function impairments (173). The hippocampal changes not only impair episodic memory but also, because of the role of the hippocampus in regulating the hypothalamus, affect HPA-axis activity (174). However, even with prospective links between self-reported stress and hippocampal measures, we cannot rule out the possibility that different brain structures make people *vulnerable* to stress or depression.

In addition to substantial evidence of links between stress, isolation, and socioeconomic conditions and brain structure in older adults, there is some evidence that such differences begin to emerge early in life. Among young children, parental SES predicts systematic differences in neurocognitive test performance (especially related to executive function and verbal fluency); structural features of the brains including both volumetric (175, 176) and prefrontal cortical thickness measures (177); and neural recruitment during problem solving (178). In a small EEG study of infants, lower parental income and occupation were associated with lower frontal gamma power (considered indicators of the brain's capacity to support attentional processes) (179). These infants were assessed at an average age of 226 days, so reverse causation is not plausible. The analyses adjusted for several possible confounders, such as birthweight and sleep, but the study is nonetheless observational and the reasons for the association are not necessarily attributable to stress.

Adoption studies, both observational and randomized, have provided critical evidence that the associations described above reflect the effects of social adversity on brain structure, rather than reverse causation. For example, in one observational study of 43 internationally adopted children, longer time with the birth family prior to being placed for adoption predicted better executive function, as did higher quality of orphanage, as rated by parents and adoption agencies (180). Internationally adopted children were also shown to have worse white matter integrity than nonadopted children, an association that attenuated with longer time in the adoptive home (181).

In a remarkable randomized trial of study of Romanian orphans aged 9 to 31 months (the Bucharest Early Intervention Project), marked baseline (prerandomization) EEG alpha and beta band power differences were observed between institutionalized and noninstitutionalized children (182). Some institutionalized children were randomly assigned to receive high-quality foster care, while others were randomly assigned to "usual care," which typically implied a longer period in the orphanage and, when adopted, foster parents without intensive supports and training from the Early Intervention Project. Outcomes depended on the age at which children were randomly assigned to the foster care. Children randomized to high-quality foster care showed substantial improvements in cognitive and developmental outcomes compared with children randomized to

"usual care" (183), and the magnitude of effects were larger for children randomized at younger ages. Among those institutionalized children randomly assigned to receive high-quality foster care before 24 months of age, differences in EEG parameters compared with noninstitutionalized children were no longer statistically significant at 8 years of age. Differences in EEG (compared with never institutionalized children) persisted for those institutionalized children randomized to "usual care" or to foster care at age greater than 24 months (184). Note that the sample is quite small and EEG indicators for children randomized to receive high-quality foster care were actually intermediate between never-institutionalized children and the "control" children randomized to remain in the institution: the key intention-to-treat analysis between children randomized to high quality foster care and children randomized to remain in the institution was not statistically significant. Studies following orphans, in particular the Bucharest Early Intervention Project, describe the consequences of extreme exposures that are likely more profoundly injurious than experiences of poverty or social isolation within the typical spectrum, but may nonetheless provide insight about how social adversity becomes embodied.

CELLULAR AGING: TELOMERES AND OXIDATIVE STRESS

Oxidative stress—defined as excess generation of free radicals relative to antioxidant defense—has been implicated in cancer development, the aging process, and neurodegenerative disorders. Chronic social stress is associated with higher levels of oxidative stress and lower levels of antioxidants (113). For example, one study found prospectively assessed education levels and occupational status were associated with high levels of oxidation as measured by concentrations of F(2) isoprostanes and gamma-glutamyltransferase, lower levels of antioxidants as measured by carotenoids; lower levels of education and occupational status also predicted greater increases in gamma-glutamyltransferase and greater decreases in carotenoids over 10 to 15 years of follow-up (185).

Similarly, a number of recent studies have suggested that leukocyte telomere length (LTL), a marker of cellular aging, is also sensitive to effects of social stress (186–189) and can provide an early indication of premature aging. Telomeres are DNA-protein complexes that cover the ends of chromosomes and promote stability of the chromosome.[3] Telomere maintenance seems to play an important role in determining longevity and may be involved in disease development. Numerous cross-sectional studies have linked shorter telomeres in human leukocytes with a variety of age-related conditions and diseases, including cardiovascular disease (193), increases in insulin resistance (194), and cancer (195). Relatedly, reduced telomerase activity has been linked with behavioral and biological risk factors for cardiovascular disease (188). Several prospective studies have examined telomere length in relation to longevity, and most but not all found that telomere length predicted earlier mortality (196–198). Much of the evidence suggesting that cellular aging,

3 Telomere length varies over time due to both lengthening and shortening activities in the cell (190). The rate of shortening in telomeres is partly controlled by the activity of telomerase, a cellular enzyme that adds telomeric repeat sequences to the chromosomal DNA ends. Telomerase activity serves not only to maintain telomere length but also to preserve healthy cell function, prolonged stem cell proliferation, and long-term immune function, independent of telomere length (191, 192).

as reflected by shorter telomeres, might be associated with organismic aging has been derived from cross-sectional studies. However the most recent work has begun to consider whether rate of change in telomere length shortening might predict mortality. For example, in a study of 236 older adults in which telomere length was measured at 2 time points, 2.5 years apart, rate of telomere length change predicted mortality (199). In this study, men with leukocyte telomere shortening over the 2.5 year follow-up were subsequently three times more likely to die from heart disease than those who maintained leukocyte telomere length over the same period. One possible implication of the findings, should they be replicated, is that telomere length changes over the short term provide a clinically useful measure of health status and risk. However, much remains to be learned about telomere dynamics and their relation to health.

Studies have considered telomere length and telomerase dynamics in relation to social adversity either to assess whether telomere shortening may be one mechanism underlying health-related effects of social adversity or to use telomere length as a predisease proxy suggesting likely health effects in the absence of manifest disease. Efforts to elucidate effects of social adversity on health in relation to race/ethnicity and telomeres provide an example of ongoing work in this area. Given disproportionately high levels of social disadvantage and poor health among blacks (200), the prevailing expectation has been that blacks would have shorter telomeres relative to whites. Surprisingly, most recent studies examining LTL suggest that on average, blacks in fact have longer telomeres (e.g., 201, 202, 203). For example, a cross-sectional analysis of 2,453 adults found significantly longer LTLs among blacks compared with whites (202), and similar results were obtained from a study of 667 adolescents (203). Most recently, a cross-sectional study of 2,599 high functioning black and white older adults considered the interrelationships between race/ethnicity and educational attainment. Older individuals with only a high school education had shorter LTLs than those with post–high school education; blacks had longer LTLs than whites regardless of educational attainment but the apparently protective effects of educational attainment were most pronounced among blacks (204). One hypothesized (but as yet unverified) explanation for surprising racial patterns in telomere length is that blacks start with longer LTLs, but rate of shortening is more rapid in blacks due to stress-related processes (201).

HUMAN METABOLOME

The metabolome is defined as the collection of downstream products of cellular metabolism, and it consists of a large array of small molecule metabolites critical for the growth and maintenance of cells, tissues, and organisms (205). Recent advances in high-throughput technologies permit simultaneous and accurate measurement of hundreds of small molecule metabolites in biological specimens like plasma ("metabolomics"). As of this writing, the most recent release of the Human Metabolome Database included over 40,000 metabolites (206). Metabolic alterations associated with disease risk can be evaluated through the use of metabolomics platforms that quantify these metabolites.

At present, metabolomics approaches are most commonly utilized for characterizing and understanding a range of disorders, with the goal of improving diagnostic tools. However, metabolomics tools have the potential for broader applications relevant to understanding physiologic mechanisms linking adversity and health. Metabolomics may provide a method for monitoring and measuring physiological changes that occur in response to social adversity. Characterizing metabolomic

consequences of such exposures may yield not only a snapshot of cellular functioning at any given time, but (if measured repeatedly over time) also provide potentially important information on a dynamic signature of biological dysregulation that is responsive to a changing environment. Some investigators have suggested that it may be possible to detect a reliable biological "signature" of stress exposure. Other investigators, however, have questioned the value added by discovery of a metabolomics profile of social adversity, and suggested that the gain in knowledge may not warrant expense of the technology. Similar to research on the epigenome and gene expression, any specific research project in this area is exploratory and high risk because our current base of knowledge is so limited. However, as with many of the "omics" technologies, the cost of processing metabolomics data will likely decline significantly. Linking metabolomics profiles with information on social risk factors may provide important mechanistic insights into the stress-related biologic alterations that are most likely to result in disease. Because metabolites are downstream of cellular transcription and post-translational processes, and closer to the actual disease outcomes of interest, metabolomic investigations may suggest not only when these alterations occur, but also whether they are durable or could be reversed. More generally, metabolomic assessment may provide insight into key biochemical pathways by which social adversity influences health, as well as a method for evaluating whether some periods of exposure are particularly sensitive or early detection of disease risk, and if interventions may successfully alter cellular function.

However, such possibilities are speculative and there is very little work to date in this area. Initial animal studies have identified substantial metabolomic changes in relation to induced chronic stress, with alterations in amino acid, carbohydrate, and lipid metabolism (e.g., 207). For example, glutamate levels were significantly increased and glutamine levels decreased in animals exposed to chronic stress (e.g., 76, 77).[4] Whether these changes will be observed in humans is yet to be determined.

Metabolomic investigations focus on characterizing and quantifying all the small molecules in complex biological samples. A related endeavor, metabonomics, includes recognition of the human microbiome (the thousands of species of microorganisms existing in a human body, including gut, mucosal tissues, and skin). The vast majority of *cells* in a human body are not, in fact, human (209). Metabonomics aims to measure the global, dynamic metabolic response of living systems to perturbations caused by environmental factors (including diet and toxins), disease processes, and the involvement of extragenomic influences, such as gut microflora. The focus is more directed toward understanding systemic changes occurring over time in complex multicellular systems (210).

This work has suggested the potentially important role of gut flora in linking environmental exposures to health. Disruption of gut microbial activity seems to be central to certain diseases including those in the gut, liver, pancreas, and brain (210). The most relevant research for social epidemiology has emerged out of observations of the high comorbidity between stress-related psychiatric symptoms (e.g., anxious behavior) and gastrointestinal disorders, including irritable bowel syndrome and inflammatory bowel disorder, leading to the proposed concept of a microbiota–gut–brain axis (211). For example, one recent study demonstrated that altering intestinal

4 Glutamine is an amino acid critical to the synthesis of proteins and neurotransmitters that can be reversibly converted to glutamate, a molecule involved in gluconeogenesis. These metabolites are involved in regulating blood sugar levels and other cell functions (208) that have been linked with downstream health outcomes.

microbiota in mice could alter anxiety-related behavior; similarly, the investigators reported that transplanting intestinal microbiota from nonanxious mice to intestines of anxious mice could reduce anxiety behavior in the initially anxious mice (212). This is consistent with a growing body of research in germ-free animals and in animals exposed to pathogenic bacterial infections, probiotic bacteria, or antibiotic drugs suggesting that gut microbiota are involved in the regulation of anxiety, mood, cognition, and pain (211). Effects may also be bidirectional, given that stress and the associated activity of the HPA axis are also known to influence the composition of the gut microbiota (213). For example, one study demonstrated that among adult rats compared with nonseparated control animals, those that had undergone maternal separation for 3 hours per day in early life, showed altered fecal microbiota composition (214). Other work has corroborated these findings in humans, demonstrating that chronic stress (measured in adulthood) affected the gut microbiota composition (for recent review, see 211). Although still nascent, this research suggests that gut microbiome composition could mediate links between social adversity and health or exacerbate effects of stressful experiences.

The findings reviewed above suggest important insights may come out of studying how the microbiome is altered by social environmental exposures. Because there are thousands of species of microorganisms in the human gut, it is impossible to study each one in isolation to work out what they do (211). However, the emerging techniques for "omics" data can be used to provide greater understanding of whether exogenous perturbations from the social environment lead to specific alterations in the microbiome that in turn influence metabolic pathways relevant to life-long health.

GENOME, EPIGENOME, AND TRANSCRIPTOME

Regulation of gene expression[5] is now thought to play an important role in physiologic responses to social adversity. Gene expression is a dynamic process: it may be up- and down-regulated throughout life; it is responsive to environmental conditions; and it varies systematically by developmental stage (17). Existing evidence implicates regulation of DNA transcription into RNA, for example via epigenetic marks on the DNA, as well as post-translational modifications of protein building blocks, as likely mechanisms by which the social environment is biologically embedded (29, 37).

EPIGENETICS

One mechanism by which social conditions are hypothesized to influence gene expression is through epigenetic modifications. Methylation and acetylation of sections of DNA can induce stable changes in gene activity without altering the underlying DNA sequence (215). These epigenetic "marks" alter access of regulatory proteins and RNA polymerase to the DNA, thus reducing transcription of DNA into RNA, and reducing production of the gene's protein product.

5 Including both transcription from DNA into RNA and translation from RNA into protein.

Although the typical direct effect of DNA hypermethylation is to down-regulate expression of the relevant gene, many proteins can up- or down-regulate the expression of *other* genes, so epigenetic marks on one gene may increase likelihood of expression of other genes. Epigenetic modifications are an intrinsic part of typical developmental processes and play a critical role in healthy human development.

Three aspects of epigenetic processes make them appealing mechanisms to explain enduring social inequalities in health. First, epigenetic patterns are largely established early in life, with episodes of rapid DNA methylation and demethylation occurring at specific developmental periods (e.g., in primordial, unfertilized germ cells, and shortly after fertilization). After these episodes, epigenetic patterns are largely preserved over the lifecourse of an individual, although there are some changes, termed "epigenetic drift" (216). The extent of epigenetic drift in humans (after early periods of rapid, large-scale demethylation and remethylation) is uncertain and an active area of research. It *is* recognized however, that because of the relative stability of epigenetic modifications, they provide a possible mechanism for a "sensitive" period exposure in the lifecourse, such that environmentally induced changes in early life may persist into adulthood. Second, epigenetic patterns in parental DNA influence epigenetic patterns in offspring DNA, although the links are not entirely straightforward. Finally, external environmental conditions also influence methylation and demethylation events, so epigenetic patterns may be a footprint of past environmental conditions of the organism and his or her parents. Although epigenetic patterns are largely stable except for brief developmental episodes, when epigenetic marks *do* change, it has been shown that the patterning is responsive to environmental conditions, such as nutrition, stress, infection, and toxins. Because germ cells for the next generation are present in the parental embryo, epigenetic modifications can also provide a mechanism for intergenerational transmission of phenotypic adaptations. Epigenetic modifications are therefore hypothesized to provide a very flexible physiologic tool that allows an organism to adopt a developmental path that is best fitted to its own specific environment (217) based on the conditions that prevail early in development.

Much of the excitement about epigenetics is based on theoretical promise, rather than empirical evidence. A handful of supporting animal models are very compelling. A highly influential set of rat studies demonstrated that epigenetic modifications mediated a link between a mother's nurturing behavior in the early postnatal period and her offspring's stress response; these stress response differences predicted disease risk throughout the offspring's adult life. Specifically, both experimentally induced and naturally occurring variations in rat mothers' licking and grooming of their pups in the period 1–6 days after birth predicted decreased methylation of a promoter for a glucocorticoid receptor gene. This promoter altered glucocorticoid receptor production in hippocampal tissue, which in turn led to enhanced glucocorticoid feedback sensitivity, less corticosterone and plasma ACTH release, and a generally muted HPA stress response profile (218, 219). The methylation patterns established in the early postnatal period were stable throughout the offspring's life such that as a result differences in maternal behavior predicted cognitive differences in their offspring even as the latter reached older ages (220). The assumption based on these studies has been that such difference in stress response would translate into poorer health in multiple domains. In a more recent mouse epigenetic study, parent generation fear conditioning associated with a particular odor led to elevations in the sensitivity of both their children and grandchildren to that odor. This transmission was

thought to be mediated by hypomethylation of the gene for the relevant olfactory receptor, manifest in the sperm of the exposed parent (221).

Study of such epigenetic processes in humans however, has proven difficult because under normal circumstances, the timing of epigenetic modifications is specific to developmental stages and may be tissue-specific. In addition, few human studies have data available from which to assess whether potential epigenetic modifications evident in childhood predict health outcomes in adulthood. As a result, despite the tremendous promise of epigenetic mechanisms to explain the influence of early life disadvantage on adult health, research is currently in early stages. For example, in a sample of 40 adult males selected from the 1958 British birth cohort for extremes of childhood SES and adult SES, methylation differences in gene promoters were more frequently evident in relation to childhood SES, although also observed for adult SES (222). The specific link with childhood SES may suggest that differences in childhood conditions leave an epigenetic footprint. Recent work in humans has also identified systematically different epigenetic profiles among individuals who have experienced significant trauma and post-traumatic stress and those who have not (e.g., 223). In a set of studies relying exclusively on brain tissue from suicide victims, proxy reports that the victim experienced child abuse were associated with methylation patterns analogous to the methylation patterns of "low licking and grooming" mice (224), suggesting that adversity in childhood may alter biological functioning in a manner related to both mental and physical health outcomes.

Thus, to date, studies suggest that exposure to adversity in the social environment may cause epigenetic changes in genes involved in the biological stress response (e.g., glucocorticoid receptor genes, CRH genes), especially early in life, when methylation patterns are more dynamic. Modifications to these stress responses may be linked to adult disease susceptibility via mechanisms described briefly above (29). Little research has directly identified which epigenetic modifications are most relevant in terms of incurring disease risk later in life; in other words, we do not know if epigenetic modifications of a handful of specific chromosomal regions is most important, or a whole-genome methylation pattern. Researchers are now pursuing both whole genome and candidate gene methylation differences as both consequences of social adversity and predictors of adult disease. Similarly, it is as yet unclear whether modifications occurring in one particular developmental period may be more potent than those that occur in another in relation to disease risk. Miller and colleagues have suggested that, given the importance of inflammatory processes to many diseases of aging, epigenetic modifications that establish proinflammatory tendencies in cells of the immune system are likely one critical pathway to increased susceptibility to a range of adverse health outcomes (29). Such models suggest that these tendencies are maintained by exaggerated responses to challenges and decreased sensitivity to inhibitory signals; the resulting chronic inflammation drives pathogenic mechanisms that contribute to initiating a range of diseases later in life. To date, other pathways have been less explored.

This promising research area, however, faces important financial and logistical challenges because characterizing epigenetic patterns is expensive and epigenetic patterns differ across tissues. Brain tissue is essentially inaccessible among living humans, but epigenetic modifications that occur in central tissues are likely to be importantly involved in the critical pathways. Most human studies currently evaluate epigenetic modifications in lymphocytes, but it is unclear whether lymphocytes provide a reasonable proxy for effects in the brain (38).

OTHER GENE EXPRESSION MECHANISMS

Social factors may also influence gene expression via other mechanisms. As described in a recent overview of the literature to date, research considering gene regulation and the social environment has most commonly considered a general measure of gene regulation captured by variation in steady-state transcript (mRNA) expression levels (17). Research in this area has been conducted largely in laboratory rodent models or captive primates (rhesus macaques), with a smaller set of studies in human populations. Research within humans has focused on identifying whether and how various forms of social stress are associated with differential gene expression across hundreds of genes (for comprehensive reviews, see 17, 106, 225, 226). Findings to date suggest that major aspects of social adversity predict patterns of gene expression (indicated by RNA levels) as measured in peripheral blood mononuclear cells (PBMCs) (17).

A general pattern of consistent up-regulation of genes involved in inflammation and adrenergic signaling has been identified and described as a conserved transcriptional response to adversity (CTRA) (226). Thus, CTRA refers to a shift in the basal pattern of transcription in leukocytes toward readiness to defend against various types of microbial exposures occurring under dangerous environmental conditions. Because these transcriptional shifts enhance wound healing and reduce likelihood of infection, they are highly adaptive in a physically threatening or uncertain environment for maintaining health and body integrity in the short term. However, this pattern of transcription can also contribute to excessive proinflammatory immune response gene expression and to inadequate antiviral immune response gene expression (226), and once this pattern is in play, it may in turn lead to impaired subsequent response to acute injury or infection (105). These transcriptional shifts are activated not only by physical threats but also by real and imagined social threats, the more commonly encountered threats in the contemporary social world (226). In adverse social circumstances where recurring stressful circumstances lead to a perception of ongoing danger but many of the threats encountered are nonphysical, CTRA may represent a maladaptive response that leads to increased risk of both inflammation-related diseases and to viral infections (due to insufficient antiviral immune response gene expression) (227).

The identification of the CTRA pattern built on early work in this area suggesting that prolonged stress decreases cell sensitivity to anti-inflammatory effects of glucocorticoids as evidenced by decreased expression of the GR NR_3C_1 gene, post-translational modification of the glucocorticoid receptor (GR) protein, increased expression of GR antagonists, and decreased activity of GR transcription cofactors (225, 228). For example, one study identified differentially expressed genes in the immune system among 14 individuals with and without chronically high levels of social isolation, using a cross-sectional, observational study design (225). Leukocyte transcriptional alterations were assessed with global gene expression profiling, and differential expression was found for 209 transcripts. Relative to the comparison group, isolated individuals exhibited increased expression of genes involved in immune activation, transcription control and cell proliferation, and decreased expression of genes involved in innate antiviral resistance, supporting antibody production, and mature B lymphocyte function. Results were unchanged after controlling for demographic factors, other established psychological risk factors, medical conditions, or behavioral risk factors and other biomedical parameters. Several other studies identified similarly altered gene expression patterns in relation to low SES and other forms of social adversity including social rejection, chronic interpersonal difficulties, and trauma exposure resulting in significant

distress (102, 229–232). Similar to the earlier study on social isolation, findings showed decreased activity in anti-inflammatory transcription control pathways but increased activity in proinflammatory pathways among individuals with lower versus higher adversity in various forms.

However, findings have not been entirely consistent across studies. For example, one recent study comparing gene expression patterning among individuals exposed to trauma and who developed post-traumatic stress disorder (PTSD) with those who did not develop PTSD failed to find evidence of overexpression of proinflammatory genes (233). Also important to note is that most studies to date have been observational and cross-sectional (234). Animal models of social behavior have provided more direct evidence that social conditions causally influence gene expression (17). For example, one study of female rhesus macaques manipulated dominance rank and found that it explained substantial variation in expression levels of inflammation-related immune genes in PBMCs (235). This and other animal studies provide additional support for a causal interpretation for effects of social adversity on altered gene expression patterns.

A variety of other challenges to this research have recently been discussed (17). A key outstanding issue is that while altered transcript levels in response to social and psychological factors may inform mechanisms of action, these transcriptional changes have not been clearly linked to disease development. This research faces other technical challenges similar to those identified above for epigenetic research: quantification of RNA levels for any given gene requires sufficient volumes of high-quality homogeneous cellular material, and human studies are often restricted to samples from accessible tissues such as subcutaneous adipose tissue, skeletal muscle, and peripheral blood mononuclear cells. Blood is the most feasibly collected biological sample in a population setting, but its use in gene expression assumes that it provides *general* information on transcription in different cells and tissues, including those more relevant for the phenotype or disease of interest. Moreover, few research groups have the technical capacity to perform this research, particularly among human populations. Technical challenges include obtaining appropriate samples, performing the assays, and implementing bioinformatic analyses needed to analyze the results. As a result, the body of available research is currently small and emerges from a small number of labs. Comprehensive independent replication is not likely to be feasible in the near-term given existing capacity in the field. With the rate of technical advances in the field, for example tools to extract RNA from dried blood spots, and suggestive findings to date, work in this area is likely to be taken up by more investigators, facilitating replication studies (17).

DIFFERENCES IN THE GENETIC CODE

The role of genetic differences in the creation of social inequalities in health has long been highly controversial, evidenced in the "nature versus nurture" debate (236). Although there has been little direct evidence for large effects of any specific genetic polymorphism on a social phenotype, twin studies suggest moderate to substantial heritability for many social factors, such as educational attainment (237). Opposition to a genetics explanation stems in part from the incorrect assumption that inequalities emerging from genetic factors cannot be eliminated. However, the recent decades of genetic research suggest two key insights that call for a reconsideration of

this debate. First, genome-wide association studies (GWAS) have produced surprisingly few confirmed genetic polymorphisms with large effects on complex disorders. For many complex conditions, such as diabetes, stroke, heart disease, or depression, all identified variants account for only a small fraction of variance, much smaller than heritability estimates from twin studies, prompting a discussion of how to explain the "missing heritability" (238). Findings from GWAS for psychiatric outcomes have been particularly underwhelming to date, but this area is advancing so rapidly the picture may look different in the near future (239). Most published GWAS have been concerned with identifying disease-associated loci, but more recently GWAS has been extended to examine variation in psychosocial phenotypes (e.g., well-being) and behavior (e.g., social bonding, educational attainment [240]). However, it has proven difficult to identify simple associations of genes with complex psychosocial phenotypes. For example, a major GWAS of education, based on a discovery sample of 101,069 individuals and a replication sample of 25,490, identified only 3 genome-wide significant polymorphisms, and these 3 in combination accounted for only about 0.022% of the variance in education (241). Although this research is just beginning, based on current evidence, there is reason to anticipate the pattern for social factors will be similar to the pattern observed for complex conditions such as cardiovascular disease: there are few or no common genetic variants with large effects. If there is a strong genetic influence on social phenotypes, it is likely due to a large number of alleles, each with tiny effects.

The second insight to prompt a reconsideration of the "nature versus nurture" dichotomization is that the importance of genetic background is fully conditional on the environmental context. Recent work has been concerned with evaluating how the social environment may exacerbate or mitigate potential adverse effects of certain genes. For example, two studies using independent data indicate that state cigarette taxes modify genetic risk of smoking behavior (242, 243) and that heritability estimates of smoking behavior differ substantially across birth cohorts (who are themselves exposed to different sociocultural norms) (244). One cannot partition a phenotype into "genetic" contributions or "environmental" contributions: a number of conditions could plausibly be described as both *entirely* genetic and also *entirely* environmental. Phenylketonuria is the paradigmatic example: the disease develops when a polymorphism in the gene for phenylalanine hydroxylase is inherited from both parents, making it impossible for the offspring to metabolize the amino acid phenylalanine. Although the disease in this sense is entirely genetic, the clinical manifestations can be prevented by dietary changes and supplementation (i.e., altering environmental conditions). Thus, study of environmental modifiers must go hand in hand with genetic research.

Major gene environment interactions may also play a role in explaining the "missing heritability," that is, the discrepancy between high heritability estimates from twin studies and the tiny fraction of phenotypic variance explained by identified polymorphisms (245). Twin studies typically estimate heritability by contrasting similarity between monozygotic twins (100% shared genetic code) versus dizygotic twins (50% shared genetic code). Any variability due to gene-gene interactions or due to interactions between genes and shared environmental factors would be allocated to the "heritable" rather than "environmental" components in such models. This occurs because if both a gene and a specific environmental factor are necessary to manifest as disease, if one monozygotic twin has both the adverse gene and the adverse environment, then the other does as well, 100% of the time. However, if one dizygotic twin has both the adverse gene and the adverse environment, there is only a 50% chance the other has both.

MITIGATION, PLASTICITY, AND REVERSIBILITY

After reviewing the many mechanisms by which social adversity leads to pathophysiologic processes, a crucial question is: can this harm be fixed? In other words, can the effects of adversity be reversed if the adversity is removed? Research to date is limited but suggests that in fact, plasticity in some domains, including brain structure and function, persists into adulthood and old age and can be triggered by environmental context (246–248). For example, in a study of the effects of postnatal maternal separation, Francis and colleagues found that environmental enrichment during the peripubertal period completely reversed effects of maternal separation on HPA and behavioral responses to stress, but there was no effect on CRF mRNA expression (249). Radley and colleagues found that effects of restraints stress in rats on apical dendritic retraction and axospinous synapse loss in the medial prefrontal cortex could be completely reversed with the removal of stress exposure (250). A small number of human studies support the conclusions from animal studies. One study examined the effects of reducing distress on inflammation in a population of cancer patients, and found that when distress was effectively reduced, levels of inflammation also decreased (251). In a study of the effects of relaxation response, individuals who engaged in relaxation response practices for a sustained period of time exhibited altered patterns of gene expression that may be associated with better health outcomes (252). Taken together this work suggests that effects of adversity may well be modifiable, but that plasticity depends on timing of mitigating exposures (in terms of developmental period), and the timing of the exposure of the initial stress. In general, the stimuli that trigger plasticity in the adult organism may differ substantially from the stimuli that trigger plasticity early in development. Moreover, while modifiability may not be achievable in all systems, relatively effective compensation may be possible.

FUTURE DIRECTIONS AND CRITICAL CHALLENGES

The past two decades have witnessed a dramatic increase in research on the physiologic mechanisms underlying social inequalities in health. This research has occurred in many different disciplines, often without optimal interdisciplinary communication, and has developed explanations at multiple physiologic levels, from DNA to behavior. We stand at a particularly exciting moment because the combination of theoretical and technical advances recently achieved should support rapid advances in our understanding within the next few years. Below we discuss some of the most promising, high-priority avenues to pursue in the next generation of research.

HUMAN SYSTEMS INTEGRATION: CAPITALIZING ON THE "OMICS"

Increasing awareness that biological systems interact has led to greater efforts within social epidemiology to characterize effects of social exposures on multiple biological systems simultaneously.

Fully integrating this vision into our research will require advances in several areas related to both measurement and analysis.

DESIGNING MEASURES THAT ADVANCE MECHANISTIC UNDERSTANDING

An early example of a multisystem measure was the introduction of the concept of allostatic load (253), first proposed by McEwen and Stellar (33) and subsequently elaborated by McEwen (27), with empirical studies following soon thereafter (254, 255). Allostatic load research has provided convincing evidence of the biological correlates of social adversity, an important first step to establishing causality. Next, to guide thinking about interventions, we will need to clearly address specific mechanisms and sequencing of physiologic events. We need greater insight into the timing of adverse exposures, the most relevant components of adversity, and potential reversibility of the effects of exposures.

Other multisystem measures, such as metabolic syndrome, have proven informative both to predict long-term disease risk and as a measure of the biological impact of the social environment across a broad range of systems. However, most indices (as they exist currently) primarily include markers of physiologic deterioration (e.g., inflammation, poor endothelial function) but do not assess *adaptive* physiological responses to stress. Understanding adaptation and plasticity will likely be important for understanding resilience and recovery in the context of social adversity (256).

Many existing measures are operationalized with arbitrary cutpoints defining a single high-risk threshold, for example based on the highest tertile or quartile in the sample distribution. In reality, nearly all biomarkers are likely associated with risk in a graded, rather than binary, fashion; nonmonotonic associations, with high disease risk associated with both low and high levels of the biomarker, are also common. Application of arbitrary and inappropriate cutpoints could theoretically introduce substantial measurement error, reduce effective statistical power, and obscure important differences in effects across the risk distribution. That said, comparative work operationalizing allostatic load using the full range of values for component markers and incorporating nonlinear risk scoring has (at least to date) provided only minimally better risk prediction than the original, relatively crude measures. The importance of imperfect measurement probably depends on how the measure is to be used (as an outcome, control variable, or predictor), and one *advantage* of combining indicators across multiple systems is that measurement errors are averaged across multiple instruments.

ANALYZING COMPLEX DATA

The huge investments in describing the human genome, transcriptome, proteome, metabolome, and human biome should also provide more opportunities to evaluate the molecular and cellular mechanisms that may be influenced by the social environment. For example, explicit evaluation of gene-environment interactions may help explain some long-standing puzzles in social epidemiology: why do some people flourish in adversity while others experience terrible health? How can

high heritability estimates for many diseases be consistent with very large social inequalities in health, if social factors are truly causal? The movement for "personalized" medicine has emphasized ways in which clinical treatment decisions might be guided by individual genetics, but social context should perhaps play an equally large role in considering clinical needs (257). Moreover, results in genomic research have largely reaffirmed the importance of modifiable environmental factors. The idea that optimal phenotypes are defined with respect to the environment has long been recognized (258) and is clearly consistent with major explanatory theories in social epidemiology, such as the "differential susceptibility to context" idea. However, social epidemiology has by and large not played a huge role in defining the gene-environment research agenda, in part because this type of gene-environment interaction has been challenging to establish (259). Despite the challenges of research at the intersection of the "omics" and social epidemiology, we anticipate these intersections will help illuminate the physiologic pathways linking social adversity and health.

Further complexity arises because biological systems and biomarkers of interest are not simply working together in additive fashion, but rather interact with one another in potentially complex ways. Anatomic systems are generally characterized by complexity that manifests itself as fractal (i.e., self-similar or branching "tree-like") patterns, such as the branching nature of the nervous, pulmonary, circulatory, and His-Purkinje systems, which promotes the rapid transfer of information or nutrients and provides excess capacity and nonlinear capability. As a result, investigators have suggested that current analytic techniques tend to rely too heavily on assumptions of linearity and have highlighted the need for analyses that can more explicitly account for nonlinearities in various systems (260). Moreover, as each system is biologically complex in its own right and works on a different timescale (210), it is challenging to integrate broadly across multiple systems while simultaneously accounting for the complexity of each system. New analytic techniques such as network analysis (knowledge-driven) or computational modeling of all system-level data (data-driven) are becoming available and make it possible to maximize "omics" data (261). Integrating the specialized knowledge and tools to conduct this research well requires technical and substantive expertise beyond the reach of most individuals working at the interface between social science and biology. Such work inevitably will require collaborative multidisciplinary teams.

Despite seemingly perpetual calls for interdisciplinary work—and the remarkable success of a handful of ambitious initiatives (262)—building, funding, and stabilizing teams with multidisciplinary expertise, or even individuals with meaningful interdisciplinary training, remains a critical challenge (263). To date, there are fewer investigators with high-level expertise in biology that are also committed to research on social processes than vice versa. As a result, a small number of individuals with expertise in a particular biological system may contribute to most studies in a given area. This is an important disadvantage, because independent replication by outside investigators is difficult or impossible. Such concerns may be mitigated as techniques for conducting "omics" research become more accessible but in the meantime should factor into assessment of the literature and findings to date.

EVIDENCE TO GUIDE ACTION

Finally, with multisystem outcomes, and complex feedback processes, there is a risk of uninformative conclusions that "everything affects everything." Additional progress will come from more

specific tests of competing hypotheses with inconsistent predictions. Currently, the evidence linking social adversity to health suggests multiple plausible mechanisms, but to date little research has delineated a clear biological sequencing of the primary mechanisms, when damage is incurred, and how modifiable the damage may be. This is an ambitious research agenda, but ultimately such work will more clearly provide a guide to appropriate action. One approach that has been suggested is to take a disease-centered perspective that "reverse engineer" adverse health outcomes into their specific biological determinants and then assesses whether and when those processes are modified by the social environment (106). Such an approach could promote insight into the mechanisms underlying how social factors influence health, provide early markers of susceptibility or disease risk, help to identify the strongest etiologic period for intervention, and provide a biological imprint of a particular exposure.

Of the many important theoretical questions, we consider the question of recovery and plasticity to be of especially high priority. Can the effects of toxic social exposures on health, once they have occurred, be reversed? If so, what conditions most effectively trigger or support such recovery and when are they most likely to be effective? To date, only a handful of studies have directly examined whether effects of toxic social exposures may be reversed. Most of the work in human populations has been done among adult patient populations (i.e., individuals who have already developed disease such as heart disease or cancer), and these studies have been inconsistently successful in mitigating effects of toxic exposures, particularly in relation to physical health (see Chapter 11; 2). Findings from the Bucharest Early Intervention Project reveal the importance of early remediation efforts. However, we anticipate that the boundaries of adult plasticity have not yet been established, and despite the importance of early developmental periods, effective strategies for remediation in adults can be found.

TRIANGULATING ACROSS MULTIPLE RESEARCH MODALITIES TO SUPPORT CAUSAL INFERENCES

Debates about the causal direction linking social adversity with health are pervasive (3, 264) because with conventional epidemiologic research approaches, it is impossible to rule out confounding or reverse causation. A key strategy is to build conclusions on research that utilizes a range of study designs since natural and intentional experiments can significantly bolster and support evidence obtained from observational studies. While each design may have its own strengths and weaknesses (see Table 14.1), triangulating across studies can provide more convincing and powerful evidence of the effect of interest. For example, laboratory-based experimental research on effects of social status and acute stress can be an invaluable complement to observational studies or studies of population level interventions. Despite their more limited generalizability, laboratory studies have the advantage of investigating biological responses under standardized conditions, with stronger capacity for causal inference. One challenge for experimental research, however, is to establish that the short-term perturbations observed in a laboratory setting correspond with accumulating, long-term health changes (50, 265). Work more formally and systematically integrating evidence across research designs—defining the assumptions underlying inferences in each study and identifying opportunities to test those assumptions using other study designs—should be a high priority. This again requires individuals trained across disciplines and collaborating with

interdisciplinary teams but also holds the best promise to accelerate research on the biological foundations of social inequalities in health.

CONCLUSION

In conclusion, understanding the biology of adversity and whether biological alterations may be modified could have profound effects on both basic research and policy approaches to developing effective strategies to improve population health. Progress on the basic science side is occurring rapidly, benefiting from the burgeoning availability of biomarkers in major cohort studies and the huge investments in "omics" technologies. As a result, many potential contributing mechanisms have been identified with current evidence pointing to clear associations between such physiologic parameters and social adversity at multiple levels, from gene expression to immune and metabolic regulation. While existing evidence is largely observational (and much is cross-sectional), findings are consistent with the hypothesis that there is indeed a causal relationship *from* adversity *to* health, rather than from health to social conditions. Establishing this causal direction is a fundamental challenge in social epidemiology, as described in several other chapters in this text, so this contribution of the physiology research is important. But it is not enough. Our current knowledge is not sufficiently specific to guide either the content or timing of interventions. Concepts of life-course timing have been inconsistently invoked in biological research and in general, alternative models of etiologic periods have not been explicitly tested against one another. The available evidence appears to support the biological plausibility of special importance of early developmental periods for some processes, but we also know that some level of biological plasticity is maintained into adulthood. A key question for future research is the range of plasticity in various biological systems that is maintained through childhood and into adulthood, and how to elicit this plasticity in the service of better health and well-being. Success in answering this and other crucial questions regarding relationships between social adversity, physiology, and health will depend on our ability to develop the interdisciplinary teams needed to (1) more clearly articulate the multisystems models associated with competing social and biological theories and (2) implement the research designs needed to test these models.

REFERENCES

1. Whalley B, Thompson DR, Taylor RS. Psychological interventions for coronary heart disease: Cochrane systematic review and meta-analysis. Int J Behav Med. 2012:1–13.
2. Berkman LF. Social epidemiology: social determinants of health in the United States: are we losing ground? Annu Rev Public Health. 2009;30:27–41.
3. Macleod J, Davey Smith G. Psychosocial factors and public health: a suitable case for treatment? J Epidemiol Community Health. 2003;57(8):565–70.
4. Smith JP. Unraveling the SES-Health Connection. In: Waite L, editor. Aging, health, and public policy: demographic and economic perspectives. New York: The Population Council: Population and Development Review Supplements; 2005. pp. 108–32.

5. Mohanan M. Causal effects of health shocks on consumption and debt: quasi-experimental evidence from bus accident injuries. Rev Econ Stat. 2013;95(2):673–81.
6. Wagstaff A, Lindelow M. Are health shocks different? Evidence from a multishock survey in Laos. Health Economics. 2013. The World Bank, Policy Research Working Paper Series: 5335, 2010.
7. Cacioppo JT. Social neuroscience: understanding the pieces fosters understanding the whole and vice versa. Am Psychol. 2002;57(11):819–30.
8. Singer BH, Ryff CD, editors. New horizons in health: an integrative approach. Washington, DC: National Academy Press; 2001.
9. Harris JR, Gruenewald T, Seeman T. An overview of biomarker research from community and population-based studies. In: Weinstein M, Vaupel JW, Wachter KW, editors. Biosocial surveys. Washington, DC: National Academies Press; 2008. pp. 96–135.
10. Matthews KA, Gallo LC, Taylor SE. Are psychosocial factors mediators of socioeconomic status and health connections? A progress report and blueprint for the future. Ann N Y Acad Sci. 2010;1186:146–73.
11. Juster RP, McEwen BS, Lupien SJ. Allostatic load biomarkers of chronic stress and impact on health and cognition. Neurosci Biobehav Rev. 2010;35(1):2–16.
12. Brulle RJ, Pellow DN. Environmental justice: human health and environmental inequalities. Annu Rev Public Health. 2006;27:103–24.
13. Crowder K, Downey L. Inter-neighborhood migration, race, and environmental hazards: modeling micro-level processes of environmental inequality. Am J Sociol. 2010;115(4):1110.
14. Health, United States, 2011. Hyattsville, MD: National Center for Health Statistics. 2012.
15. Harper S, Lynch J. Trends in socioeconomic inequalities in adult health behaviors among US states, 1990–2004. Public Health Rep. 2007;122(2):177.
16. Glass TA, Bandeen-Roche K, McAtee M, Bolla K, Todd AC, Schwartz BS. Neighborhood psychosocial hazards and the association of cumulative lead dose with cognitive function in older adults. Am J Epidemiol. 2009;169(6):683–92.
17. Tung J, Gilad Y. Social environmental effects on gene regulation. Cell Mol Life Sci. 2013.
18. Marmot MG, Shipley MJ, Hemingway H, Head J, Brunner EJ. Biological and behavioural explanations of social inequalities in coronary heart disease: the Whitehall II study. Diabetologia. 2008;51(11):1980–8.
19. Stringhini S, Sabia S, Shipley M, Brunner E, Nabi H, Kivimäki M, et al. Association of socioeconomic position with health behaviors and mortality. JAMA. 2010;303(12):1159–66.
20. Marmot MG, Rose G, Shipley M, Hamilton PJ. Employment grade and coronary heart disease in British civil servants. J Epidemiol Community Health. 1978;32(4):244–9.
21. Banks J, Marmot M, Oldfield Z, Smith JP. Disease and disadvantage in the United States and in England. JAMA. 2006;295(17):2037–45.
22. McEwen BS. Central effects of stress hormones in health and disease: understanding the protective and damaging effects of stress and stress mediators. Eur J Pharmacol. 2008;583(2–3):174–85.
23. Selye H. Stress and disease. Science. 1955;122(3171):625–31.
24. Ewbank DC. Biomarkers in social science research on health and aging: a review of theory and practice. In: Weinstein M, Vaupel JW, Wachter KW, editors. Biosocial surveys. Washington, DC: National Academies Press; 2008. pp. 156–71.
25. Segerstrom SC, Miller GE. Psychological stress and the human immune system: a meta-analytic study of 30 years of inquiry. Psychol Bull. 2004;130(4):601–30.
26. Cohen S, Janicki-Deverts D, Miller GE. Psychological stress and disease. JAMA. 2007;298(14):1685–7.

27. McEwen BS. Protective and damaging effects of stress mediators. N Engl J Med. 1998;338(3):171–9.
28. Relman AS, Angell M. Resolved: psychosocial interventions can improve clinical outcomes in organic disease (con). Psychosom Med. 2002;64(4):558–63.
29. Miller GE, Chen E, Parker KJ. Psychological stress in childhood and susceptibility to the chronic diseases of aging: moving toward a model of behavioral and biological mechanisms. Psychol Bull. 2011;137(6):959–97.
30. Cannon WB. The wisdom of the body. New York: Norton; 1932.
31. Kitano H. Towards a theory of biological robustness. Mol Syst Biol. 2007;3:137.
32. Sterling P, Eyer J. Allostasis: a new paradigm to explain arousal pathology. In: Fisher S, Reason J, editors. Handbook of life stress, cognition and health. New York: Wiley & Sons; 1988. pp. 631–51.
33. McEwen BS, Stellar E. Stress and the individual: mechanisms leading to disease. Arch Intern Med. 1993;153(18):2093–101.
34. Seeman T, Epel E, Gruenewald T, Karlamangla A, McEwen BS. Socio-economic differentials in peripheral biology: cumulative allostatic load. Ann N Y Acad Sci. 2010;1186:223–39.
35. Gunnar M, Quevedo K. The neurobiology of stress and development. Annu Rev Psychol. 2007;58: 145–73.
36. Geronimus AT, Bound J, Waidmann TA, Colen CG, Steffick D. Inequality in life expectancy, functional status, and active life expectancy across selected black and white populations in the United States. Demography. 2001;38(2):227–51.
37. Hertzman C. Putting the concept of biological embedding in historical perspective. Proc Natl Acad Sci U S A. 2012;109 Suppl 2:17160–7.
38. Rutter M. Achievements and challenges in the biology of environmental effects. Proc Natl Acad Sci U S A. 2012;109(Suppl 2):17149–53.
39. Hales CN, Barker DJ. The thrifty phenotype hypothesis. Br Med Bull. 2001;60:5–20.
40. Gluckman PD, Hanson MA. Developmental origins of disease paradigm: a mechanistic and evolutionary perspective. Pediatr Res. 2004;56(3):311–7.
41. Barker DJ. Fetal origins of coronary heart disease. BMJ. 1995;311(6998):171–4.
42. Roseboom TJ, van der Meulen JHP, Osmond C, Barker DJP, Ravelli ACJ, Schroeder-Tanka JM, et al. Coronary heart disease after prenatal exposure to the Dutch famine, 1944–45. BMJ. 2000;84(6):595.
43. Obradovic J, Bush NR, Stamperdahl J, Adler NE, Boyce WT. Biological sensitivity to context: the interactive effects of stress reactivity and family adversity on socioemotional behavior and school readiness. Child Dev. 2010;81(1):270–89.
44. Shonkoff JP, Garner AS. The lifelong effects of early childhood adversity and toxic stress. Pediatrics. 2012;129(1):e232–46.
45. Eriksson PS, Perfilieva E, Bjork-Eriksson T, Alborn AM, Nordborg C, Peterson DA, et al. Neurogenesis in the adult human hippocampus. Nat Med. 1998;4(11):1313–7.
46. Elbert T, Rockstroh B. Reorganization of human cerebral cortex: the range of changes following use and injury. Neuroscientist. 2004;10(2):129–41.
47. Gervain J, Vines BW, Chen LM, Seo RJ, Hensch TK, Werker JF, et al. Valproate reopens critical-period learning of absolute pitch. FNSYS. 2013;7:102.
48. Kolb B, Gibb R. Brain plasticity and behaviour in the developing brain. J Can Acad Child Adolesc Psychiatry. 2011;20(4):265–76.
49. Steptoe A, Marmot M. The role of psychobiological pathways in socio-economic inequalities in cardiovascular disease risk. Eur Heart J. 2002;23(1):13–25.

50. Chida Y, Steptoe A. Greater cardiovascular responses to laboratory mental stress are associated with poor subsequent cardiovascular risk status: a meta-analysis of prospective evidence. Hypertension. 2010;55(4):1026–32.
51. Black PH, Garbutt LD. Stress, inflammation and cardiovascular disease. J Psychosom Res. 2002;52(1): 1–23.
52. Danese A, Caspi A, Williams B, Ambler A, Sugden K, Mika J, et al. Biological embedding of stress through inflammation processes in childhood. Mol Psychiatry. 2011;16(3):244–6.
53. Appleton AA, Buka SL, McCormick MC, Koenen KC, Loucks EB, Gilman SE, et al. Emotional functioning at age 7 years is associated with C-reactive protein in middle adulthood. Psychosom Med. 2011;73(4): 295–303.
54. Ellis BJ, Jackson JJ, Boyce WT. The stress response systems: universality and adaptive individual differences. Dev Rev. 2006;26(2):175–212.
55. Blumenshine P, Egerter S, Barclay CJ, Cubbin C, Braveman PA. Socioeconomic disparities in adverse birth outcomes: a systematic review. Am J Prev Med. 2010;39(3):263–72.
56. Barker D. Developmental origins of chronic disease. Public Health. 2012;126(3):185–9.
57. Almond D, Currie J. Killing me softly: the fetal origins hypothesis. J Econ Perspect. 2011;25(3):153–72.
58. Barker DJ, Osmond C. Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. Lancet. 1986;327(8489):1077–81.
59. Risnes KR, Vatten LJ, Baker JL, Jameson K, Sovio U, Kajantie E, et al. Birthweight and mortality in adulthood: a systematic review and meta-analysis. Int J Epidemiol. 2011;40(3):647–61.
60. Rinaudo P, Wang E. Fetal programming and metabolic syndrome. Annu Rev Physiol. 2012;74:107–30.
61. Glymour MM, Benjamin EJ, Kosheleva A, Gilsanz P, Curtis LH, Patton KK. Early life predictors of atrial fibrillation-related mortality: Evidence from the health and retirement study. Health Place. 2013;21:133–9.
62. Van der Kooy K, van Hout H, Marwijk H, Marten H, Stehouwer C, Beekman A. Depression and the risk for cardiovascular diseases: systematic review and meta analysis. Int J Geriatr Psych. 2007;22(7):613–26.
63. Pan A, Sun Q, Okereke OI, Rexrode KM, Hu FB. Depression and risk of stroke morbidity and mortality. JAMA. 2011;306(11):1241–9.
64. Manrique-Garcia E, Sidorchuk A, Hallqvist J, Moradi T. Socioeconomic position and incidence of acute myocardial infarction: a meta-analysis. J Epidemiol Community Health. 2011;65(4):301–9.
65. Avendaño M, Kunst AE, Huisman M, Lenthe FV, Bopp M, C B, et al. Educational level and stroke mortality: a comparison of 10 European populations during the 1990s. Stroke. 2004;35:432–7.
66. Avendano M, Kawachi I, Van Lenthe F, Boshuizen HC, Mackenbach JP, Van den Bos GAM, et al. Socioeconomic status and stroke incidence in the US elderly - the role of risk factors in the EPESE study. Stroke. 2006;37(6):1368–73.
67. Galobardes B, Smith GD, Lynch JW. Systematic review of the influence of childhood socioeconomic circumstances on risk for cardiovascular disease in adulthood. Ann Epidemiol. 2006;16(2):91–104.
68. Glymour MM, Avendano M, Berkman LF. Is the "stroke belt" worn from childhood? Risk of first stroke and state of residence in childhood and adulthood. Stroke. 2007;38(9):2415–21.
69. Howard VJ, McClure LA, Glymour MM, Cunningham SA, Kleindorfer DO, Crowe M, et al. Effect of duration and age at exposure to the Stroke Belt on incident stroke in adulthood. Neurology. 2013;80(18):1655–61.
70. Conen D, Tedrow UB, Cook NR, Buring JE, Albert CM. Birth weight is a significant risk factor for incident atrial fibrillation. Circulation. 2010;122(8):764.
71. Nabi H, Chastang JF, Lefevre T, Dugravot A, Melchior M, Marmot MG, et al. Trajectories of depressive episodes and hypertension over 24 years: the Whitehall II prospective cohort study. Hypertension. 2011;57(4):710–6.

72. Kavanagh A, Bentley RJ, Turrell G, Shaw J, Dunstan D, Subramanian SV. Socioeconomic position, gender, health behaviours and biomarkers of cardiovascular disease and diabetes. Soc Sci Med. 2010;71(6):1150–60.

73. Thurston RC, Matthews KA. Racial and socioeconomic disparities in arterial stiffness and intima media thickness among adolescents. Soc Sci Med. 2009;68(5):807–13.

74. Steptoe A, Feldman PJ, Kunz S, Owen N, Willemsen G, Marmot M. Stress responsivity and socioeconomic status: a mechanism for increased cardiovascular disease risk? Eur Heart J. 2002;23(22):1757–63.

75. Mendelson T, Thurston RC, Kubzansky LD. Affective and cardiovascular effects of experimentally-induced social status. Health Psychol. 2008;27(4):482–9.

76. Depke M, Fusch G, Domanska G, Geffers R, Volker U, Schuett C, et al. Hypermetabolic syndrome as a consequence of repeated psychological stress in mice. Endocrinology. 2008;149(6):2714–23.

77. Li ZY, Zheng XY, Gao XX, Zhou YZ, Sun HF, Zhang LZ, et al. Study of plasma metabolic profiling and biomarkers of chronic unpredictable mild stress rats based on gas chromatography/mass spectrometry. RCM. 2010;24(24):3539–46.

78. Wang X, Zhao T, Qiu Y, Su M, Jiang T, Zhou M, et al. Metabonomics approach to understanding acute and chronic stress in rat models. J Proteome Res. 2009;8(5):2511–8.

79. Ni Y, Su M, Lin J, Wang X, Qiu Y, Zhao A, et al. Metabolic profiling reveals disorder of amino acid metabolism in four brain regions from a rat model of chronic unpredictable mild stress. FEBS Lett. 2008;582(17):2627–36.

80. Zhang WY, Liu S, Li HD, Cai HL. Chronic unpredictable mild stress affects myocardial metabolic profiling of SD rats. J Pharm Biomed Anal. 2012;70:534–8.

81. Cheng S, Rhee EP, Larson MG, Lewis GD, McCabe EL, Shen D, et al. Metabolite profiling identifies pathways associated with metabolic risk in humans. Circulation. 2012;125(18):2222–31.

82. Barzilaym JI, Freedland ES. Inflammation and its relationship to insulin resistance, type 2 diabetes mellitus, and endothelial dysfunction. Metab Syndr Relat Disord. 2003;1(1):55–67.

83. Panza F, Solfrizzi V, Logroscino G, Maggi S, Santamato A, Seripa D, et al. Current epidemiological approaches to the metabolic-cognitive syndrome. J Alzheimers Dis. 2012;30(Suppl 2):S31–75.

84. Elliott P, Chambers JC, Zhang W, Clarke R, Hopewell JC, Peden JF, et al. Genetic loci associated with C-reactive protein levels and risk of coronary heart disease. JAMA. 2009;302(1):37–48.

85. Ridker PM, Rifai N, Rose L, Buring JE, Cook NR. Comparison of C-reactive protein and low-density lipoprotein cholesterol levels in the prediction of first cardiovascular events. N Eng J Med. 2002;347(20):1557–65.

86. Aiello AE, Kaplan GA. Socioeconomic position and inflammatory and immune biomarkers of cardiovascular disease: applications to the panel study of income dynamics. Biodemography Soc Biol. 2009;55(2):178–205.

87. Galkina E, Ley K. Immune and inflammatory mechanisms of atherosclerosis. Annu Rev Immunol. 2009;27(1):165–97.

88. Wu J, Lanier LL. Natural killer cells and cancer. Adv Cancer Res. 2003;90:127–56.

89. Whitman SC, Ramsamy TA. Participatory role of natural killer and natural killer T cells in atherosclerosis: lessons learned from in vivo mouse studies. Can J Physiol Pharm. 2006;84(1):67–75.

90. Danesh J, Collins R, Appleby P, Peto R. Association of fibrinogen, C-reactive protein, albumin, or leukocyte count with coronary heart disease: meta-analyses of prospective studies. JAMA. 1998;279(18):1477–82.

91. Cohen S. The Pittsburgh common cold studies: psychosocial predictors of susceptibility to respiratory infectious illness. Int J Behav Med. 2005;12(3):123–31.

92. Danese A, Pariante CM, Caspi A, Taylor A, Poulton R. Childhood maltreatment predicts adult inflammation in a life-course study. Proc Natl Acad Sci. U.S.A. 2007;104(4):1319–24.

93. Pollitt RA, Rose KM, Kaufman JS. Evaluating the evidence for models of life course socioeconomic factors and cardiovascular outcomes: a systematic review. BMC Public Health. 2005;5(1):7.
94. Brunner E, Marmot M, Canner R, Beksinska M, Davey Smith G, O'Brien J. Childhood social circumstances and psychosocial and behavioural factors as determinants of plasma fibrinogen. Lancet. 1996;347(9007):1008–13.
95. Kiecolt-Glaser JK, Gouin JP, Weng NP, Malarkey WB, Beversdorf DQ, Glaser R. Childhood adversity heightens the impact of later-life caregiving stress on telomere length and inflammation. Psychosom Med. 2011;73(1):16–22.
96. Copeland WE, Shanahan L, Worthman C, Angold A, Costello EJ. Cumulative depression episodes predict later C-reactive protein levels: a prospective analysis. Biol Psychiatry. 2012;71(1):15–21.
97. Slopen N, Kubzansky LD, Koenen KC. Internalizing and externalizing behaviors predict elevated inflammatory biomarkers in childhood. Psychoneuroendocrinology. 2014;38(12):2854–62.
98. Steptoe A, Hamer M, Chida Y. The effects of acute psychological stress on circulating inflammatory factors in humans: a review and meta-analysis. Brain Behav Immun. 2007;21(7):901–12.
99. Slopen N, Kubzansky LD, McLaughlin KA, Koenen KC. Childhood adversity and inflammatory processes in youth: A prospective study. Psychoneuroendocrinology. 2012.
100. Slopen N, Koenen KC, Kubzansky LD. Childhood adversity and immune and inflammatory biomarkers associated with cardiovascular risk in youth: a systematic review. Brain Behav Immun. 2012;26(2):239–50.
101. Miller GE, Chen E. Harsh family climate in early life presages the emergence of a proinflammatory phenotype in adolescence. Psychol Sci. 2010;21(6):848–56.
102. Miller GE, Rohleder N, Cole SW. Chronic interpersonal stress predicts activation of pro- and anti-inflammatory signaling pathways 6 months later. Psychosom Med. 2009;71(1):57–62.
103. Kiecolt-Glaser JK, Marucha PT, Malarkey WB, Mercado AM, Glaser R. Slowing of wound healing by psychological stress. Lancet. 1995;346:1194–6.
104. Walburn J, Vedhara K, Hankins M, Rixon L, Weinman J. Psychological stress and wound healing in humans: a systematic review and meta-analysis. J Psychosom Res. 2009;67(3):253–71.
105. Godbout JP, Glaser R. Stress-induced immune dysregulation: implications for wound healing, infectious disease and cancer. J Neuroimmune Pharm. 2006;1(4):421–7.
106. Miller G, Chen E, Cole SW. Health psychology: developing biologically plausible models linking the social world and physical health. Annu Rev Psychol. 2009;60:501–24.
107. Miller GE, Chen E, Zhou ES. If it goes up, must it come down? Chronic stress and the hypothalamic-pituitary-adrenocortical axis in humans. Psychol Bull. 2007;133(1):25–45.
108. Dickerson SS, Kemeny ME. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. Psychol Bull. 2004;130(3):355–91.
109. Adam EK, Hawkley LC, Kudielka BM, Cacioppo JT. Day-to-day dynamics of experience—cortisol associations in a population-based sample of older adults. Proc Natl Acad Sci U S A. 2006;103(45):17058–63.
110. Raison CL, Miller AH. When not enough is too much: the role of insufficient glucocorticoid signaling in the pathophysiology of stress-related disorders. Am J Psychiat. 2003;160(9):1554–65.
111. Kroboth PD, Salek FS, Pittenger AL, Fabian TJ, Frye RF. DHEA and DHEA-S: a review. J Clin Pharmacol. 1999;39(4):327–48.
112. Morgan CA 3rd, Southwick S, Hazlett G, Rasmusson A, Hoyt G, Zimolo Z, et al. Relationships among plasma dehydroepiandrosterone sulfate and cortisol levels, symptoms of dissociation, and objective performance in humans exposed to acute stress. Arch Gen Psychiatry. 2004;61(8):819–25.

113. Bauer ME, Jeckel CM, Luz C. The role of stress factors during aging of the immune system. Ann N Y Acad Sci. 2009;1153:139–52.

114. Smith GD, Ben-Shlomo Y, Beswick A, Yarnell J, Lightman S, Elwood P. Cortisol, testosterone, and coronary heart disease: prospective evidence from the Caerphilly study. Circulation. 2005;112(3):332–40.

115. Thaker PH, Sood AK. Neuroendocrine influences on cancer biology. Semin Cancer Biol. 2008;18(3): 164–70.

116. van Raalte DH, Ouwens DM, Diamant M. Novel insights into glucocorticoid-mediated diabetogenic effects: towards expansion of therapeutic options? Eur J Clin Invest. 2009;39(2):81–93.

117. Chrousos GP. The role of stress and the hypothalamic-pituitary-adrenal axis in the pathogenesis of the metabolic syndrome: neuro-endocrine and target tissue-related causes. Int J Obes Relat Metab Disord. 2000;24(Suppl 2):S50–5.

118. Taylor SE, Lerner JS, Sage RM, Lehman BJ, Seeman TE. Early environment, emotions, responses to stress, and health. J Pers. 2004;72(6):1365–93.

119. Cohen S, Schwartz JE, Epel E, Kirschbaum C, Sidney S, Seeman T. Socioeconomic status, race, and diurnal cortisol decline in the Coronary Artery Risk Development in Young Adults (CARDIA) Study. Psychosom Med. 2006;68(1):41–50.

120. Lupien SJ, King S, Meaney MJ, McEwen BS. Child's stress hormone levels correlate with mother's socioeconomic status and depressive state. Biol Psychiatry. 2000;48(10):976–80.

121. Dowd JB, Simanek AM, Aiello AE. Socio-economic status, cortisol and allostatic load: a review of the literature. Int J Epidemiol. 2009;38(5):1297–309.

122. Kroboth PD, Salek FS, Pittenger AL, Fabian TJ, Frye RF. DHEA and DHEA-S: a review. J Clin Pharmacol. 1999;39(4):327–48.

123. Maninger N, Wolkowitz OM, Reus VI, Epel ES, Mellon SH. Neurobiological and neuropsychiatric effects of dehydroepiandrosterone (DHEA) and DHEA sulfate (DHEAS). Front Neuroendocrin. 2009;30(1):65–91.

124. Khelil MB, Tegethoff M, Meinlschmidt G, Jamey C, Ludes B, Raul J-S. Simultaneous measurement of endogenous cortisol, cortisone, dehydroepiandrosterone, and dehydroepiandrosterone sulfate in nails by use of UPLC–MS–MS. Anal Bioanal Chem. 2011;401(4):1153–62.

125. Kubzansky LD, Adler GK. Aldosterone: a forgotten mediator of the relationship between psychological stress and heart disease. Neurosci Biobehav Rev. 2010;34(1):80–6.

126. Emanuele E, Geroldi D, Minoretti P, Coen E, Politi P. Increased plasma aldosterone in patients with clinical depression. Arch Med Res. 2005;36(5):544–8.

127. Murck H, Held K, Ziegenbein M, Kunzel H, Koch K, Steiger A. The renin-angiotensin-aldosterone system in patients with depression compared to controls--a sleep endocrine study. BMC Psychiatry. 2003;3:15.

128. Walton KG, Pugh ND, Gelderloos P, Macrae P. Stress reduction and preventing hypertension: preliminary support for a psychoneuroendocrine mechanism. J Altern Complement Med. 1995;1(3):263–83.

129. Whaley-Connell A, Johnson MS, Sowers JR. Aldosterone: role in the cardiometabolic syndrome and resistant hypertension. Prog Cardiovasc Dis. 2010;52(5):401–9.

130. Fox CS, Gona P, Larson MG, Selhub J, Tofler G, Hwang SJ, et al. A multi-marker approach to predict incident CKD and microalbuminuria. J Am Soc Nephrol. 2010;21(12):2143–9.

131. Tomaschitz A, Pilz S, Ritz E, Grammer T, Drechsler C, Boehm BO, et al. Association of plasma aldosterone with cardiovascular mortality in patients with low estimated GFR: the Ludwigshafen Risk and Cardiovascular Health (LURIC) Study. Am J Kidney Dis. 2011;57(3):403–14.

132. Koal T, Schmiederer D, Pham-Tuan H, Rohring C, Rauh M. Standardized LC-MS/MS based steroid hormone profile-analysis. J Steroid Biochem Mol Biol. 2012;129(3–5):129–38.

133. Bigger JT, Jr., Fleiss JL, Steinman RC, Rolnitzky LM, Kleiger RE, Rottman JN. Frequency domain measures of heart period variability and mortality after myocardial infarction. Circulation. 1992;85(1):164–71.

134. Kleiger RE, Miller JP, Bigger JT Jr., Moss AJ. Decreased heart rate variability and its association with increased mortality after acute myocardial infarction. Am J Cardiol. 1987;59(4):256–62.

135. Molgaard H, Sorensen KE, Bjerregaard P. Attenuated 24-h heart rate variability in apparently healthy subjects, subsequently suffering sudden cardiac death. Clin Auton Res. 1991;1:223–33.

136. Thayer JF, Lane RD. The role of vagal function in the risk for cardiovascular disease and mortality. Biol Psychol. 2007;74(2):224–42.

137. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Circulation. 1996;93(5):1043–65.

138. van Ravenswaaij-Arts CMA, Kollee LAA, Hopman JCW, Stoelinga GBA, van Geijn HP. Heart rate variability. Ann Intern Med. 1993;118:436–47.

139. Chandola T, Britton A, Brunner E, Hemingway H, Malik M, Kumari M, et al. Work stress and coronary heart disease: what are the mechanisms? Eur Heart J. 2008;29(5):640–8.

140. Lucini D, Di Fede G, Parati G, Pagani M. Impact of chronic psychosocial stress on autonomic cardiovascular regulation in otherwise healthy subjects. Hypertension. 2005;46(5):1201–6.

141. Michels N, Sioen I, Clays E, De Buyzere M, Ahrens W, Huybrechts I, et al. Children's heart rate variability as stress indicator: Association with reported stress and cortisol. Biol Psychol. 2013;94(2):433–40.

142. Kemp AH, Quintana DS. The relationship between mental and physical health: Insights from the study of heart rate variability. Int J Psychophysiol. 2013;89(3):288–96.

143. Kemp AH, Quintana DS, Gray MA, Felmingham KL, Brown K, Gatt JM. Impact of depression and antidepressant treatment on heart rate variability: a review and meta-analysis. Biol Psychiatry. 2010;67(11):1067–74.

144. Suls J. Anger and the heart: perspectives on cardiac risk, mechanisms and interventions. Prog Cardiovasc Dis. 2013;55(6):538–47.

145. Thayer JF, Yamamoto SS, Brosschot JF. The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors. Int J Cardiol. 2010;141(2):122–31.

146. Hemingway H, Shipley M, Brunner E, Britton A, Malik M, Marmot M. Does autonomic function link social position to coronary risk? The Whitehall II study. Circulation. 2005;111(23):3071–7.

147. Sloan RP, Huang MH, Sidney S, Liu K, Williams OD, Seeman T. Socioeconomic status and health: is parasympathetic nervous system activity an intervening mechanism? Int J Epidemiol. 2005;34(2):309–15.

148. Sloan RP, Huang MH, McCreath H, Sidney S, Liu K, Dale Williams O, et al. Cardiac autonomic control and the effects of age, race, and sex: the CARDIA study. Auton Neurosci-Basic. 2008;139(1–2):78–85.

149. Thayer JF, Brosschot JF. Psychosomatics and psychopathology: looking up and down from the brain. Psychoneuroendocrinology. 2005;30(10):1050–8.

150. Buffalari DM, Grace AA. Chronic cold stress increases excitatory effects of norepinephrine on spontaneous and evoked activity of basolateral amygdala neurons. Int J Neuropsychoph. 2009;12(1):95–107.

151. Swiergiel AH, Leskov IL, Dunn AJ. Effects of chronic and acute stressors and CRF on depression-like behavior in mice. Behav Brain Res. 2008;186(1):32–40.

152. Swinny JD, O'Farrell E, Bingham BC, Piel DA, Valentino RJ, Beck SG. Neonatal rearing conditions distinctly shape locus coeruleus neuronal activity, dendritic arborization, and sensitivity to corticotrophin-releasing factor. Int J Neuropsychoph. 2010;13(4):515–25.

153. Beauchaine TP, Neuhaus E, Zalewski M, Crowell SE, Potapova N. The effects of allostatic load on neural systems subserving motivation, mood regulation, and social affiliation. Dev Psychopathol. 2011;23(4):975–99.

154. Goldstein DS, Kopin IJ. Adrenomedullary, adrenocortical, and sympathoneural responses to stressors: a meta-analysis. Endocr Regul. 2008;42(4):111–9.

155. Janicki-Deverts D, Cohen S, Adler NE, Schwartz JE, Matthews KA, Seeman TE. Socioeconomic status is related to urinary catecholamines in the Coronary Artery Risk Development in Young Adults (CARDIA) study. Psychosom Med. 2007;69(6):514–20.

156. Cohen S, Doyle WJ, Baum A. Socioeconomic status is associated with stress hormones. Psychosom Med. 2006;68(3):414–20.

157. Evans GW, English K. The environment of poverty: multiple stressor exposure, psychophysiological stress, and socioemotional adjustment. Child Dev. 2002;73(4):1238–48.

158. Dowd JB, Goldman N. Do biomarkers of stress mediate the relation between socioeconomic status and health? J Epidemiol Community Health. 2006;60(7):633–9.

159. Gersten O, Dow WD, Rosero-Bixby L. Stressors over the life course and neuroendocrine system dysregulation in Costa Rica. J Aging Health. 2010;22(6):748–71.

160. Forsman L, Lundberg U. Consistency in catecholamine and cortisol excretion in males and females. Pharmacol Biochem Behav. 1982;17(3):555–62.

161. Peaston RT, Lennard TW, Lai LC. Overnight excretion of urinary catecholamines and metabolites in the detection of pheochromocytoma. J Clin Endocrinol Metab. 1996;81(4):1378–84.

162. Fuchs E, Tumbar T, Guasch G. Socializing with the neighbors: stem cells and their niche. Cell. 2004;116(6):769–78.

163. Powell ND, Sloan EK, Bailey MT, Arevalo JMG, Miller GE, Chen E, et al. Social stress up-regulates inflammatory gene expression in the leukocyte transcriptome via β-adrenergic induction of myelopoiesis. Proc Natl Acad Sci U S A. 2013;110(41):16574–9.

164. Christoffel DJ, Golden SA, Russo SJ. Structural and synaptic plasticity in stress-related disorders. Rev Neurosci. 2011;22(5):535–49.

165. Harvey P, Pruessner J, Czechowska Y, Lepage M. Individual differences in trait anhedonia: a structural and functional magnetic resonance imaging study in non-clinical subjects. Mol Psychiatr. 2007;12(8):767–75.

166. Pizzagalli DA, Holmes AJ, Dillon DG, Goetz EL, Birk JL, Bogdan R, et al. Reduced caudate and nucleus accumbens response to rewards in unmedicated individuals with major depressive disorder. Am J Psychiat. 2009;166(6):702–10.

167. Butterworth P, Cherbuin N, Sachdev P, Anstey KJ. The association between financial hardship and amygdala and hippocampal volumes: results from the PATH through life project. Soc Cogn Affect Neur. 2012;7(5):548–56.

168. Gianaros PJ, Marsland AL, Sheu LK, Erickson KI, Verstynen TD. Inflammatory pathways link socioeconomic inequalities to white matter architecture. Cereb Cortex. 2013;23(9):2058–71.

169. Gianaros PJ, Horenstein JA, Hariri AR, Sheu LK, Manuck SB, Matthews KA, et al. Potential neural embedding of parental social standing. Soc Cogn Affect Neur. 2008;3(2):91.

170. Gianaros PJ, Jennings JR, Sheu LK, Greer PJ, Kuller LH, Matthews KA. Prospective reports of chronic life stress predict decreased grey matter volume in the hippocampus. Neuroimage. 2007;35(2):795–803.

171. Ganzel BL, Kim P, Glover GH, Temple E. Resilience after 9/11: multimodal neuroimaging evidence for stress-related change in the healthy adult brain. Neuroimage. 2008;40(2):788–95.

172. McEwen BS, Morrison JH. The brain on stress: vulnerability and plasticity of the prefrontal cortex over the life course. Neuron. 2013;79(1):16–29.

173. Shah AK, Mullainathan S, Shafir E. Some consequences of having too little. Science. 2012;338(6107):682–5.

174. McEwen BS, Gianaros PJ. Stress-and allostasis-induced brain plasticity. Ann Rev Med. 2011;62:431–45.

175. Hanson JL, Adluru N, Chung MK, Alexander AL, Davidson RJ, Pollak SD. Early neglect is associated with alterations in white matter integrity and cognitive functioning. Child Dev. 2013;84(5):1566–78.

176. Noble KG, Houston SM, Kan E, Sowell ER. Neural correlates of socioeconomic status in the developing human brain. Developmental Sci. 2012;15(4):516–27.

177. Lawson GM, Duda JT, Avants BB, Wu J, Farah MJ. Associations between children's socioeconomic status and prefrontal cortical thickness. Developmental Sci. 2013;16(5):641–52.

178. Hackman DA, Farah MJ, Meaney MJ. Socioeconomic status and the brain: mechanistic insights from human and animal research. Nat Rev Neurosci. 2010;11(9):651–9.

179. Tomalski P, Moore DG, Ribeiro H, Axelsson EL, Murphy E, Karmiloff-Smith A, et al. Socioeconomic status and functional brain development: associations in early infancy. Developmental Sci. 2013;16(5):676–87.

180. Hostinar CE, Stellern SA, Schaefer C, Carlson SM, Gunnar MR. Associations between early life adversity and executive function in children adopted internationally from orphanages. Proc Natl Acad Sci U S A. 2012;109(Suppl 2):17208–12.

181. Kumar A, Behen ME, Singsoonsud P, Veenstra AL, Wolfe-Christensen C, Helder E, et al. Microstructural abnormalities in language and limbic pathways in orphanage-reared children a diffusion tensor imaging study. J Child Neurol. 2014;29(3):318–25.

182. Saby JN, Marshall PJ. The utility of EEG band power analysis in the study of infancy and early childhood. Dev Neuropsychol. 2012;37(3):253–73.

183. Nelson CA, III, Zeanah CH, Fox NA, Marshall PJ, Smyke AT, Guthrie D. Cognitive recovery in socially deprived young children: the Bucharest early intervention project. Science. 2007;318(5858):1937–40.

184. Vanderwert RE, Marshall PJ, Nelson III CA, Zeanah CH, Fox NA. Timing of intervention affects brain electrical activity in children exposed to severe psychosocial neglect. PLoS One. 2010;5(7):e11415.

185. Janicki-Deverts D, Cohen S, Matthews KA, Gross MD, Jacobs DR, Jr. Socioeconomic status, antioxidant micronutrients, and correlates of oxidative damage: the Coronary Artery Risk Development in Young Adults (CARDIA) study. Psychosom Med. 2009;71(5):541–8.

186. Cherkas LF, Aviv A, Valdes AM, Hunkin JL, Gardner JP, Surdulescu GL, et al. The effects of social status on biological aging as measured by white-blood-cell telomere length. Aging Cell. 2006;5:361–5.

187. Epel ES, Blackburn EH, Lin J, Dhabhar FS, Adler NE, Morrow JD, et al. Accelerated telomere shortening in response to life stress. Proc Natl Acad Sci U S A. 2004;101(49):17312–5.

188. Epel ES, Lin J, Wilhelm FH, Wolkowitz OM, Cawthon R, Adler NE, et al. Cell aging in relation to stress arousal and cardiovascular disease risk factors. Psychoneuroendocrinology. 2006;31(3):277–87.

189. Simon NM, Smoller JW, McNamara KL, Maser RS, Zalta AK, Pollack MH, et al. Telomere shortening and mood disorders: preliminary support for a chronic stress model of accelerated aging. Biol Psychiatry. 2006;60:432–5.

190. Blackburn EH. Telomere states and cell fates. Nature. 2000;408(6808):53–6.

191. Kim M, Xu L, Blackburn EH. Catalytically active human telomerase mutants with allele-specific biological properties. Exp Cell Res. 2003;288(2):277–87.

192. Marrone A, Walne A, Dokal I. Dyskeratosis congenita: telomerase, telomeres and anticipation. Curr Opin Genet Dev. 2005;15(3):249–57.

193. Serrano AL, Andres V. Telomeres and cardiovascular disease: does size matter? Circ Res. 2004;94(5):575–84.

194. Gardner JP, Li S, Srinivasan SR, Chen W, Kimura M, Lu X, et al. Rise in insulin resistance is associated with escalated telomere attrition. Circulation. 2005;111(17):2171–7.

195. Aubert G, Lansdorp PM. Telomeres and aging. Physiol Rev. 2008;88(2):557–79.

196. Cawthon RM, Smith KR, O'Brien E, Sivatchenko A, Kerber RA. Association between telomere length in blood and mortality in people aged 60 years or older. Lancet. 2003;361(9355):393–5.

197. Harris SE, Deary IJ, MacIntyre A, Lamb KJ, Radhakrishnan K, Starr JM, et al. The association between telomere length, physical health, cognitive ageing, and mortality in non-demented older people. Neurosci Lett. 2006;406(3):260–4.

198. Martin-Ruiz CM, Gussekloo J, van Heemst D, von Zglinicki T, Westendorp RG. Telomere length in white blood cells is not associated with morbidity or mortality in the oldest old: a population-based study. Aging Cell. 2005;4(6):287–90.

199. Epel ES, Merkin SS, Cawthon R, Blackburn EH, Adler NE, Pletcher MJ, et al. The rate of leukocyte telomere shortening predicts mortality from cardiovascular disease in elderly men. Aging (Albany NY). 2009;1(1):81–8.

200. Williams DR, Yu Y, Jackson JS, Anderson NB. Racial differences in physical and mental health: socioeconomic status, stress, and discrimination. J Health Psychol. 1997;2:335–51.

201. Chen W, Kimura M, Kim S, Cao X, Srinivasan SR, Berenson GS, et al. Longitudinal versus cross-sectional evaluations of leukocyte telomere length dynamics: age-dependent telomere shortening is the rule. J Gerontol A-Biol. 2011;66(3):312–9.

202. Hunt SC, Chen W, Gardner JP, Kimura M, Srinivasan SR, Eckfeldt JH, et al. Leukocyte telomeres are longer in African Americans than in whites: the National Heart, Lung, and Blood Institute Family Heart Study and the Bogalusa Heart Study. Aging Cell. 2008;7:451–8.

203. Zhu H, Wang X, Gutin B, Davis CL, Keeton D, Thomas J, et al. Leukocyte telomere length in healthy Caucasian and African-American adolescents: relationships with race, sex, adiposity, adipokines, and physical activity. J Pediatr. 2011;158(2):215–20.

204. Adler N, Pantell M, O'Donovan A, Blackburn E, Cawthon R, Koster A, et al. Educational attainment and late life telomere length in the Health, Aging and Body Composition Study. Brain Behav Immun. 2012.

205. Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB. Metabolomics by numbers: acquiring and understanding global metabolite data. Trends Biotechnol. 2004;22(5):245–52.

206. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0: the human metabolome database in 2013. Nucleic Acids Res. 2013;41(D1):D801-D7.

207. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, et al. Metabolite profiles and the risk of developing diabetes. Nat Med. 2011;17(4):448–53.

208. Newsholme P, Procopio J, Lima MM, Pithon-Curi TC, Curi R. Glutamine and glutamate: their central role in cell metabolism and function. Cell Biochem Funct. 2003;21(1):1–9.

209. Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. Host-bacterial mutualism in the human intestine. Science. 2005;307(5717):1915–20.

210. Nicholson JK, Lindon JC. Systems biology: metabonomics. Nature. 2008;455(7216):1054–6.

211. Cryan JF, Dinan TG. Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. Nat Rev Neurosci. 2012;13(10):701–12.

212. Bercik P, Denou E, Collins J, Jackson W, Lu J, Jury J, et al. The intestinal microbiota affect central levels of brain-derived neurotropic factor and behavior in mice. Gastroenterology. 2011;141(2):599–609, e1–3.

213. Tannock GW, Savage DC. Influences of dietary and environmental stress on microbial populations in the murine gastrointestinal tract. Infect Immun. 1974;9(3):591–8.

214. O'Mahony SM, Marchesi JR, Scully P, Codling C, Ceolho AM, Quigley EM, et al. Early life stress alters behavior, immunity, and microbiota in rats: implications for irritable bowel syndrome and psychiatric illnesses. Biol Psychiatry. 2009;65(3):263–7.

215. Champagne FA. Epigenetic influence of social experiences across the lifespan. Dev Psychobiol. 2010;52(4):299–311.

216. Feil R, Fraga MF. Epigenetics and the environment: emerging patterns and implications. Nat Rev Genet. 2012;13(2):97–109.

217. Burdge GC, Lillycrop KA. Nutrition, epigenetics, and developmental plasticity: implications for understanding human disease. Annu Rev Nutr. 2010;30(1):315–39.

218. Champagne F, Meaney MJ. Like mother, like daughter: evidence for non-genomic transmission of parental behavior and stress responsivity. Prog Brain Res. 2001;133:287–302.

219. Szyf M, Bick J. DNA methylation: a mechanism for embedding early life experiences in the genome. Child Dev. 2013;84(1):49–57.

220. Meaney MJ, Aitken DH, Bhatnagar S, Sapolsky RM. Postnatal handling attenuates certain neuroendocrine, anatomical, and cognitive dysfunctions associated with aging in female rats. Neurobiol Aging. 1991;12(1):31–8.

221. Dias BG, Ressler KJ. Parental olfactory experience influences behavior and neural structure in subsequent generations. Nat Neurosci. 2013.

222. Borghol N, Suderman M, McArdle W, Racine A, Hallett M, Pembrey M, et al. Associations with early-life socio-economic position in adult DNA methylation. Int J Epidemiol. 2012;41(1):62–74.

223. Uddin M, Aiello AE, Wildman DE, Koenen KC, Pawelec G, de Los Santos R, et al. Epigenetic and immune function profiles associated with posttraumatic stress disorder. Proc Natl Acad Sci U S A. 2010;107(20):9470–5.

224. McGowan PO, Sasaki A, D'Alessio AC, Dymov S, Labonté B, Szyf M, et al. Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. Nat Neurosci. 2009;12(3):342–8.

225. Cole SW, Hawkley LC, Arevalo JM, Sung CY, Rose RM, Cacioppo JT. Social regulation of gene expression in human leukocytes. Genome Biol. 2007;8(9):R189.

226. Slavich GM, Cole SW. The emerging field of human social genomics. Clin Psychol Sci. 2013;1(3):331–48.

227. Cole SW, Hawkley LC, Arevalo JM, Cacioppo JT. Transcript origin analysis identifies antigen-presenting cells as primary targets of socially regulated gene expression in leukocytes. Proc Natl Acad Sci U S A. 2011;108(7):3080–5.

228. Pace TW, Hu F, Miller AH. Cytokine-effects on glucocorticoid receptor function: relevance to glucocorticoid resistance and the pathophysiology and treatment of major depression. Brain Behav Immun. 2007;21(1):9–19.

229. Chen E, Miller GE, Kobor MS, Cole SW. Maternal warmth buffers the effects of low early-life socioeconomic status on pro-inflammatory signaling in adulthood. Mol Psychiatry. 2011;16(7):729–37.

230. Miller GE, Chen E, Fok AK, Walker H, Lim A, Nicholls EF, et al. Low early-life social class leaves a biological residue manifested by decreased glucocorticoid and increased proinflammatory signaling. Proc Natl Acad Sci U S A. 2009;106(34):14716–21.

231. Murphy ML, Slavich GM, Rohleder N, Miller GE. Targeted rejection triggers differential pro- and anti-inflammatory gene expression in adolescents as a function of social status. Clin Psychol Sci. 2013;1(1):30–40.

232. O'Donovan A, Sun B, Cole S, Rempel H, Lenoci M, Pulliam L, et al. Transcriptional control of monocyte gene expression in post-traumatic stress disorder. Dis Markers. 2011;30(2–3):123–32.

233. Neylan TC, Sun B, Rempel H, Ross J, Lenoci M, O'Donovan A, et al. Suppressed monocyte gene expression profile in men versus women with PTSD. Brain Behav Immun. 2011;25(3):524–31.

234. Adler N, Bush NR, Pantell MS. Rigor, vigor, and the study of health disparities. Proc Natl Acad Sci U S A. 2012;109 Suppl 2:17154–9.

235. Tung J, Barreiro LB, Johnson ZP, Hansen KD, Michopoulos V, Toufexis D, et al. Social environment is associated with gene regulatory variation in the rhesus macaque immune system. Proc Natl Acad Sci U S A. 2012;109(17):6490–5.

236. Gottfredson LS. What if the hereditarian hypothesis is true? Psychol Public Pol L. 2005;11(2):311–9.

237. Bouchard TJ, McGue M. Genetic and environmental influences on human psychological differences. J Neurobiol. 2003;54(1):4–45.

238. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747–53.

239. Kendler KS. What psychiatric genetics has taught us about the nature of psychiatric illness and what is left to learn. Mol Psychiatr. 2013;18(10):1058–66.

240. Social Science Genetic Association Consortium. 2013 [cited 2013 December 30, 2013]; Available from: http://www.ssgac.org/.

241. Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW, et al. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science. 2013;340(6139):1467–71.

242. Boardman JD. State-level moderation of genetic tendencies to smoke. Am J Public Health. 2009;99(3):480–6.

243. Fletcher JM. Why have tobacco control policies stalled? Using genetic moderation to examine policy impacts. PLoS One. 2012;7(12):e50576.

244. Boardman JD, Blalock CL, Pampel FC. Trends in the genetic influences on smoking. J Health Soc Behav. 2010;51(1):108–23.

245. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: genetic interactions create phantom heritability. Proc Natl Acad Sci U S A. 2012;109(4):1193–8.

246. Merabet LB, Hamilton R, Schlaug G, Swisher JD, Kiriakopoulos ET, Pitskel NB, et al. Rapid and Reversible Recruitment of Early Visual Cortex for Touch. PLoS ONE. 2008;3(8).

247. Carlson MC, Erickson KI, Kramer AF, Voss MW, Bolea N, Mielke M, et al. Evidence for neurocognitive plasticity in at-risk older adults: the Experience Corps program. J Gerontol A Biol Sci Med Sci. 2009;64(12):1275–82.

248. Yang JL, Hou CL, Ma N, Liu J, Zhang Y, Zhou JS, et al. Enriched environment treatment restores impaired hippocampal synaptic plasticity and cognitive deficits induced by prenatal chronic stress. Neurobiol Learn Mem. 2007;87(2):257–63.

249. Francis DD, Diorio J, Plotsky PM, Meaney MJ. Environmental enrichment reverses the effects of maternal separation on stress reactivity. J Neurosci. 2002;22(18):7840–3.

250. Radley J, Rocher A, Janssen W, Hof P, McEwen B, Morrison J. Reversibility of apical dendritic retraction in the rat medial prefrontal cortex following repeated stress. Exp Neurol. 2005;196(1):199–203.

251. Thornton LM, Andersen BL, Schuler TA, Carson WE 3rd. A psychological intervention reduces inflammatory markers by alleviating depressive symptoms: secondary analysis of a randomized controlled trial. Psychosom Med. 2009;71(7):715–24.

252. Dusek JA, Otu HH, Wohlhueter AL, Bhasin M, Zerbini LF, Joseph MG, et al. Genomic counter-stress changes induced by the relaxation response. PLoS One. 2008;3(7):e2576.

253. Crimmins EM, Seeman T. Integrating biology into demographic research on health and aging (with a focus on the MacArthur Study of Successful Aging). In: Finch CE, Vaupel JW, Kinsella K, editors. Cells and surveys: should biological measures be included in social science research? Washington, DC: National Research Council; 2001. pp. 9–41.

254. Seeman TE, Singer B, Wilkinson C, McEwen B. Exploring a new concept of cumulative biological risk: allostatic load and its health consequences. Proc Natl Acad Sci U S A. 2001;98(8):4770–5.

255. Seeman TE, Singer BH, Rowe JW, Horwitz RI, McEwen BS. The price of adaptation: allostatic load and its health consequences: MacArthur Studies of Successful Aging. Arch Intern Med. 1997;I57:2259–68.

256. King KE, Morenoff JD, House JS. Neighborhood context and social disparities in cumulative biological risk factors. Psychosom Med. 2011;73(7):572–9.

257. Patton KK, Glymour MM. In anticipation of grief using insights from social epidemiology to improve quality of care. Circulation. 2013;128(25):2725–8.

258. Lewontin RC. The triple helix: gene, organism, and environment. Cambridge, MA: Harvard University Press; 2000.

259. Duncan LE, Keller MC. A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. Am J Psychiat. 2011;168(10):1041.

260. Rozanski A, Kubzansky LD. Psychologic functioning and physical health: a paradigm of flexibility. Psychosom Med. 2005;67(Suppl 1):S47–53.

261. Glymour MM, Osypuk TL, Rehkopf DH. Invited commentary: off-roading with social epidemiology—exploration, causation, translation. Am J Epidemiol. 2013;178(6):858–63.

262. Adler NE, Stewart J. Using team science to address health disparities: MacArthur network as case example. Ann N Y Acad Sci. 2010;1186(1):252–60.

263. Robert Wood Johnson Foundation. "Robert Wood Johnson Health and Society Scholars" Program Results Report. Princeton, NJ: Robert Wood Johnson Foundation; 2008 [cited 2014 1/11/2014]; Available from: http://www.rwjf.org/content/dam/farm/reports/program_results_reports/2011/rwjf400967.

264. Macleod J, Smith GD, Heslop P, Metcalfe C, Carroll D, Hart C. Are the effects of psychosocial exposures attributable to confounding? Evidence from a prospective observational study on psychological stress and mortality. J Epidemiol Community Health. 2001;55(12):878–84.

265. Chida Y, Hamer M. Chronic psychosocial factors and acute physiological responses to laboratory-induced stress in healthy populations: a quantitative review of 30 years of investigations. Psychol Bull. 2008;134(6):829–85.

266. Muennig P, Schweinhart L, Montie J, Neidell M. Effects of a prekindergarten educational intervention on adult health: 37-year follow-up results of a randomized controlled trial. Am J Public Health. 2009;99(8):1431–7.

267. Mendelson T, Thurston RC, Kubzansky LD. Affective and cardiovascular effects of experimentally-induced social status. Health Psychol. 2008;27(4):482.

268. Allen K, Blascovich J, Mendes WB. Cardiovascular reactivity and the presence of pets, friends, and spouses: the truth about cats and dogs. Psychosom Med. 2002;64(5):727.

269. Meaney MJ. Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. Annu Rev Neurosci. 2001;24(1):1161–92.

270. Lleras-Muney A. The relationship between education and adult mortality in the US. Rev Econ Stud. 2005;72(1):189–221.

271. Catalano R, Bruckner T, Anderson E, Gould JB. Fetal death sex ratios: a test of the economic stress hypothesis. Int J Epidemiol. 2005;34(4):944–8.

272. Kubzansky LD, Kawachi I, Sparrow D. Socioeconomic status, hostility, and risk factor clustering in the Normative Aging Study: any help from the concept of allostatic load? Ann Behav Med. 1999;21(4):330–8.

273. Seeman TE, Crimmins E, Huang M-H, Singer B, Bucur A, Gruenewald T, et al. Cumulative biological risk and socio-economic differences in mortality: MacArthur studies of successful aging. Soc Sci Med. 2004;58(10):1985–97.

CHAPTER 15

FROM SCIENCE TO POLICY

Michael Marmot and Jessica Allen

INTRODUCTION

In the first edition of this book, I began this chapter by posing a series of questions about emerging biomedical research on people and animals and their potential relevance for government policies (1). I set out research evidence showing close associations between social status, economic conditions, and health, and described the existence of clear social class gradients in many causes of mortality and for life expectancy. I described how these health gradients are observable in most countries and populations and wrote about a developing approach to explaining how and why such gradients and associations between status and life expectancy and health occur. The way circumstances affect different life stages differently was introduced, and I briefly explored intergenerational transmission of factors related to health.

I began to sketch out preliminary answers to my starter questions, demonstrating effects on population health distributions of psychosocial factors, which in turn are influenced by political, social, and economic conditions. I argued that such clear associations, and emerging explanatory research, clearly showed that international, national, and local policies should be concerned with impact on health equity. We had a reasonably clear understanding of the science, and through our social determinants framework we began to think through the implications for international, national, and local policies and strategies.

To us, it is remarkable that so much has happened in the years between that exploratory chapter and this second edition chapter. In the space of thirteen years we have seen rapid development in both science and policy development on social determinants of health (SDH). The evidence base, as laid out in the chapters of this book, demonstrates the impact of social determinants on health and health distribution. Social determinants concepts and discourse seem firmly established in academic public health, but also in many policy quarters. Many of us work closely with national and local governments and a myriad of other local, national, and international organizations across sectors, advising on appropriate strategies, interventions, and policies to improve health equity through action on the SDH.

Much has been achieved in a short time, but there are also powerful pulls away from the social determinants approach. The economic crisis has taken focus and action away, just when we need *more*, not less, focus and action on the social and economic drivers of health. The powerful discourse around individual responsibility (and accountability) for health too often leads some to apportion blame to the unhealthy. Money is cut; policies that may worsen inequity or at best ignore impact are introduced. Importantly, too, public awareness about the extent and causes of health inequity remains underdeveloped and, partly because of this, political attention to the issue is frequently short-lived.

In this chapter we document some of the developments in the social determinants approach since the original chapter, including descriptions of the evidence-based commissions we have been involved in, which have drawn on the most relevant and up-to-date evidence base and have moved thinking along. We describe some of the ways the approach has been implemented across the world, drawing on examples, and we describe some of the political and economic challenges to the approaches we have advocated. Finally, we explore potential future developments and opportunities for sustaining and developing momentum and activity to tackle gross health inequalities through action on social determinants of health.

THE SOCIAL DETERMINANTS OF HEALTH APPROACH: BUILDING THE EVIDENCE AND DEVELOPING POLICIES

THE COMMISSION ON THE SOCIAL DETERMINANTS OF HEALTH

The Commission on the Social Determinants of Health (CSDH) was convened by the WHO in 2005, and Michael Marmot was asked to chair the commission (2). It seemed to the WHO that the approaches and construction of evidence on SDH had reached a point that demanded that the WHO provide an important platform to bring together the evidence on the causes of profound health inequalities across the world. This was a significant and innovative development. First, we would have a real understanding of global levels of health inequity between and within countries. Second, the CSDH represented a unique opportunity to collect and assess global evidence on social, economic, and political factors that influence health in countries at all stages of development. And third, the commission was asked to make practical proposals to tackle and reduce health inequity across the world.

It seemed remarkable, and inspiring, that the WHO, primarily an organization advising on disease control or development of healthcare services across the world, would take on the profoundly social, cultural, political, and economic challenges of SDH. It was a significant prioritization of health inequity for the WHO, and for member states, and was testament to the development of such a strong evidence base on SDH. All of us involved in the commission were strong advocates for the importance of providing proposals for action based on clear, strong evidence. All the proposals made in the CSDH report are underpinned by evidence collected and analyzed by leading academics and practitioners from around the world.

We set up knowledge networks to allow us to assess the evidence in the range of areas related to the SDH. For two years or more the knowledge networks worked to build and analyze the evidence base; their outputs are contained in the remarkable knowledge network reports that were the basis of the final report. We appointed 19 commissioners with substantial experience of government, academia, and civil society, all committed to furthering social justice—they were, and indeed remain, an inspiring group, many continuing to lead the health equity movement in their fields. We met 10 times during the course of the commission, and we all became increasingly strong, well-informed, and vocal advocates. This high level of activity helped lead to increasing recognition from the WHO of the unfair distribution of health and health outcomes, and to acceptance that this distribution was the result of unfair social, political, and economic arrangements. The commission showed, too, that dramatic social class gradients in health are clearly present within most countries across the world.

Although some of the commissioners had substantial experience in health, and others in politics, and still others in social development, few came with a worked-out approach to SDH. The secretariat, with the chair, spent considerable effort developing a conceptual framework drawing on our own work as well as that of many others. This conceptual approach, adopted by the commission, provided a framework for the analysis and the proposals for action (Figure 15.1).

In order to understand the routes to poor health and inequities, the evidence clearly showed that we should begin with analysis of the socioeconomic and political contexts, the nature of society. These broad global and national contexts frame and shape the social position of individuals within society—manifested in levels of education, occupation, and income, and related to gender, ethnicity, and race. These wider societal-level processes influence individuals' exposure to health-damaging and health-promoting conditions and affect vulnerabilities and levels of resilience (3).

Underlying the CSDH framework is the importance of empowerment: material, psychosocial, and political. The relationship between empowerment in many spheres of life and health status has

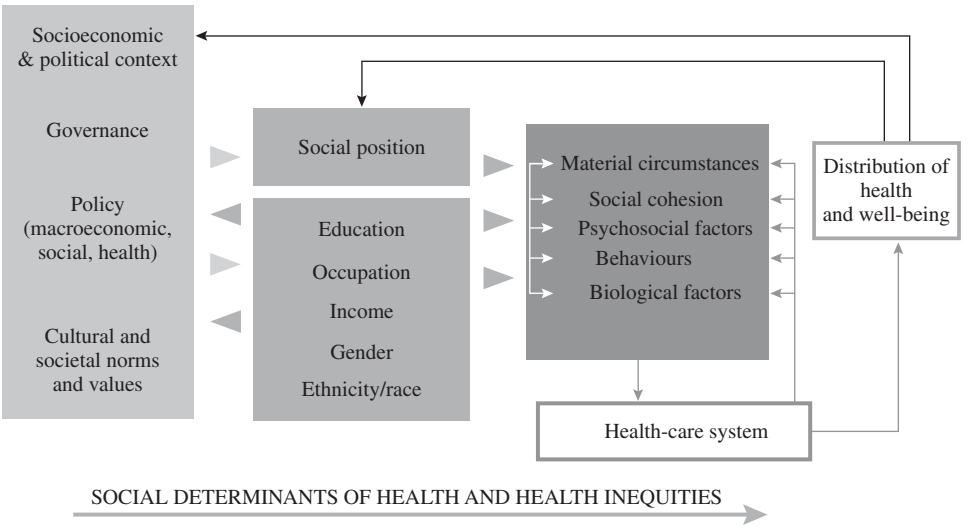


FIGURE 15.1: The conceptual approach of the CSDH.

been widely explored, and the findings appear consistent: having control over one's life leads to better health. And having control over one's life is powerfully related to social, economic, political, and cultural spheres. The evidence collected during the life of the commission led us to conclude that health inequities were determined by the conditions in which people are born, grow, live, work, and age, and by the inequities in power, money, and resources that give rise to these conditions of daily life. Successful action to tackle health inequity must be directed to these broad spheres.

This marked a clear departure from analyses that conclude that health inequity is the result of lack of access to healthcare facilities. While the ambition for universal access to healthcare must be a priority, even its successful realization will not significantly reduce health inequities; such is the significance of social and economic factors and empowerment. The CSDH demonstrated and asked that organizations committed to reducing health inequity look outside their own immediate healthcare sphere. This is challenging. In order to get the requisite approaches joined up across all sectors of society, the CSDH suggested that health ministers take the lead across government; they must make the case and make common causes with their colleagues in other departments. They must find ways of embedding health equity in all policies and demonstrate the health equity effects of policies.

There has been opposition to this cross-government approach; some sectors accuse those who try to achieve health in all policies of "health imperialism" and reject collaborations on shared agendas. We hear of governments whose siloed approach becomes more entrenched, not less. From within and outside health sectors there are strong pushes to focus on individual behaviors and healthcare sectors. Even where there is a willingness to work across governments on shared platforms of equity, the mechanisms and process of government are often set against the necessary cross-government approaches and instead foster siloed activities in departmental divisions. Sometimes the proposed mechanisms for introducing health in all policies are too mechanistic, a tick box assessment exercise that loses focus on equity. The principle of assessing policy impact on equity can become lost and faces objections from those with impact fatigue—those who are asked to assess policy impacts on wide range of criteria.

The CSDH argued, however, that improving health and health equity should be a priority for governments: health after all is a priority for the populations they serve. It also demonstrated that tackling health inequity through the SDH would produce many equity benefits in other sectors of society.

In many ways, the impact of the CSDH has been remarkable, and much activity has followed. In 2008, at the launch of the Commission Report, *Closing the Gap in a Generation*, Margaret Chan, the director general of the WHO, welcomed the focus on social justice in the CSDH report:

> The Commission on Social Determinants of Health. . . responded to a situation in which the gaps, within and between countries, in income levels, opportunities, health status, life expectancy and access to care are greater than at any time in recent history. . . In the final analysis, the distribution of health within a population is a matter of fairness in the way economic and social policies are designed. By showing how social factors directly shape health outcomes and explain inequities, the report challenged health programmes and policies to tackle the leading causes of ill-health at their roots, even when these causes lie beyond the direct control of the health sector. (4)

In 2011, following the adoption of the CSDH recommendation that there should be a global conference, the government of Brazil hosted the First World Conference on Social Determinants of Health. This gathering of representatives of 126 member states, civil society, international organizations and academics provided an opportunity to do more to galvanize support, prioritize action, and respond to the CSDH's call for social justice as a route to a fairer distribution of health. Held in October 2011, the Rio summit had as its goals to report on progress since the CSDH and to stimulate further global and national action on SDH and health equity. The conference culminated with member states' acceptance of the Rio Political Declaration on Social Determinants of Health (5). This acceptance expresses global political commitment for the implementation of an SDH approach to reduce health inequities and to achieve other global priorities. It is hoped that this commitment will help to build momentum for the development of national action plans and strategies. The International Federation of Medical Students' Associations pointed out that the Rio declaration refrained from endorsing the CSDH recommendation to tackle inequities in power, money, and resources. Such a recommendation was a step too far for some of the signatories to the declaration.

Following the CSDH, many individual countries have taken action on the social determinants, including Brazil, Peru, Chile, and India (see boxes below); below we discuss English responses, and the 2013 European Review, both inspired by the CSDH.

United Nations Development Programme (UNDP), in collaboration with Michael Marmot, is developing an SDH approach to NCDs.

Pan American Health Organization (PAHO) has prioritized SDH and health equity. Social determinants of health are on the agenda for all WHO Regions. WHO organized and ran the Rio Summit in 2011, which resulted in commitments to take forward action and review developments.

Chile—Ministry of Health review on how its policies fit the CSDH recommendations.

Argentina—appointed vice minister of health with responsibility for health equity.

Brazil—implemented Brazilian National Commission on Social Determinants of Health and co-organized World Conference on Social Determinants of Health.

Costa Rica—implemented whole-of-government approach to tackling health equity.

South Australia (Australia)—initiative from WHO and South Australia: the Adelaide Statement on Health in All Policies (6).

The Asia Pacific Network of the Global Action for Health Equity Network (AP-HealthGAEN)—regional collective progressing the health equity agenda (14).

The Malmö Commission (2010–2013)—The commissioned assembled evidence and proposed strategies for reducing health inequalities and improving the long-term living conditions for the citizens of Malmö. The approach was inspired by the CSDH.

Swedish Association of Local Authorities and Regions (SALAR) (2013)—together with 20 local authorities, county councils, and regions, has identified five recommendations and 23 measures that contribute to strengthening social sustainability and reducing health inequalities. An important point of departure for this work was the WHO report *Closing the Gap in a Generation*.

Alberta (Canada)—Provincial Government is actively developing an SDH program. There is ongoing interest in an SDH approach in work with indigenous people's health in Arctic Canada.

New Zealand—Development of an SDH approach in public health.

Peru—The Mayor of Lima's Health Strategy is influenced by the CSDH.

Initiatives in India

In India, there is some interest in creating a network for SDH. This would serve to prioritize action on the SDH and allow the collaboration and cross-sector activity required for effective action.

Other examples include:

Civil Society

Self-Employed Women's Association (SEWA)—an organization and movement representing poor, self-employed women workers (15).

Government initiatives

- rural employment guarantee scheme;
- food security bill;
- consideration of restructuring the Integrated Child Development Services;
- healthcare expenditure to rise from 1.2% to 3% of GDP related to action on the SDH;
- plans to extend coverage of social security for informal workers;
- extending the right to education;
- plans to improve housing and basic infrastructure for the urban and rural poor.

Anticorruption

- Widespread demonstrations called for strong anticorruption legislation following Anna Hazare's hunger strike in April 2011 when government talks broke down.
- Government has accepted Hazare's revisions to the Jan Lokpal Bill, a proposal to establish an independent anticorruption body (16).

Source: Mirai Chatterjee SEWA, personal communication (7).

THE MARMOT REVIEW

In 2008, the English government responded to the CSDH call for national governments to produce their own plans for action, and commissioned a review of health inequities in England, chaired by Michael Marmot. The English government was responding to evidence of persistent and potentially widening health inequities in England and requested evidence-based proposals for action at the national level. The review process was underpinned by nine task groups led by academics, who, as with the CSDH, provided evidence-based proposals for action across the social determinants in England (8).

The English Review, built on the conceptual approach of the CSDH, was adapted for national context (Figure 15.2).

There was a clear focus on lifecourse, based on evidence demonstrating how inequity begins very early in life—even prenatally—and accumulates across the lifecourse, resulting in unequal distributions of health and length of life.

The review was also reporting at a time of economic austerity and governmental cuts, and the proposals made were expected to demonstrate cost efficacy. In order to expose the economic as well as the human cost of health inequity, the review commissioned analysis of the costs of health inequity to the Treasury and public purse. Based on calculations the review reported that inequality in illness accounted for productivity losses of £31–33 billion per year, lost taxes and higher welfare payments in the range of £20–32 billion per year, and additional healthcare costs associated with well in excess of £5.5 billion. To further demonstrate the likely costs to the treasury, the review demonstrated the financial impact of raising the pension age if nothing were done to reduce

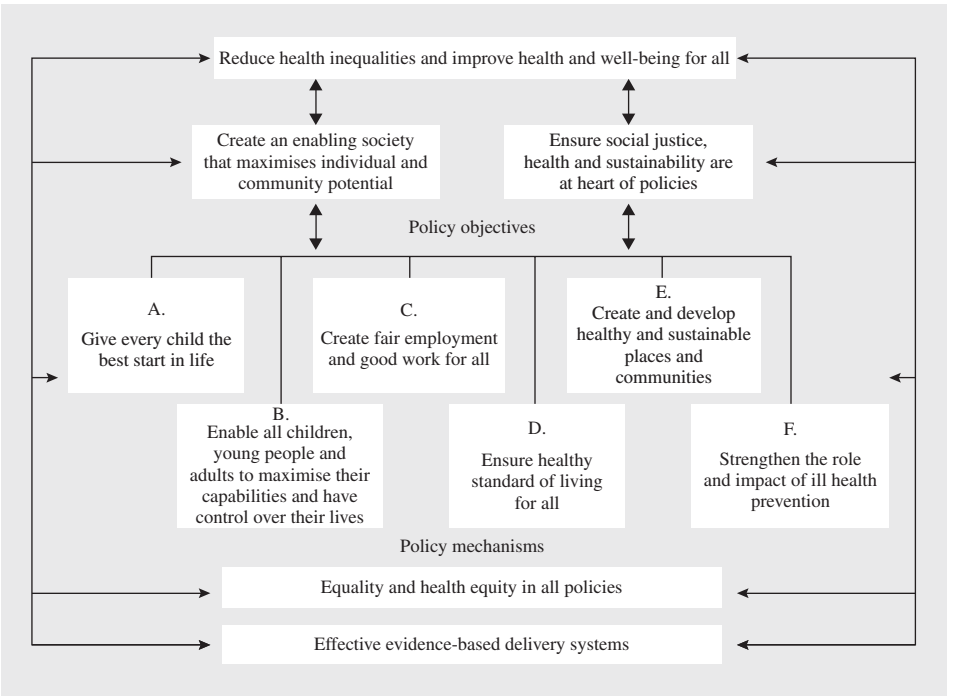


FIGURE 15.2: The Marmot review conceptual approach.

health inequities across the social gradient. Figure 15.3 depicts the English social class gradient in life expectancy and disability-free life expectancy based on levels of neighborhood deprivation in all neighborhoods in England. We have drawn a line on this graph at 68 years, the pensionable age to which England is moving. With the level of disability shown, more than three-quarters of the English population do not have disability-free life expectancy as far as the age of 68. If the population is expected to keep working until 68, it is essential to take action to improve health proportionately across the gradient.

The report of the English Review, *Fair Society, Healthy Lives*, made proposals for action under the six priority policy areas, and advocated that policies be proportionate to need across the gradient, and universal when appropriate (9). The review also provided a monitoring system to be used at the national level. While the report was commissioned by a Labour government in 2008, and reported to them in February 2010, the recommendations were accepted and supported by the incoming Conservative-led Coalition government in May 2010; cross-party political support has been important in gaining acceptance of the approaches and has ensured a level of prioritization of health equity.

In 2011, the government issued a public health white paper with an SDH focus, putting reduction of health inequalities at the center of its strategy (10). Much was based on, or a direct response to, *Fair Society, Healthy Lives.* The Public Health Outcomes Framework, outlined in the white paper, contains a tranche of indicators under SDH, along with other public health indicators, directly influenced by *Fair Society, Healthy Lives*; the new national public health agency Public Health England is based around the approach advocated in the report. In 2012, legislation introduced a set of new health inequality duties throughout the English health system, effective from

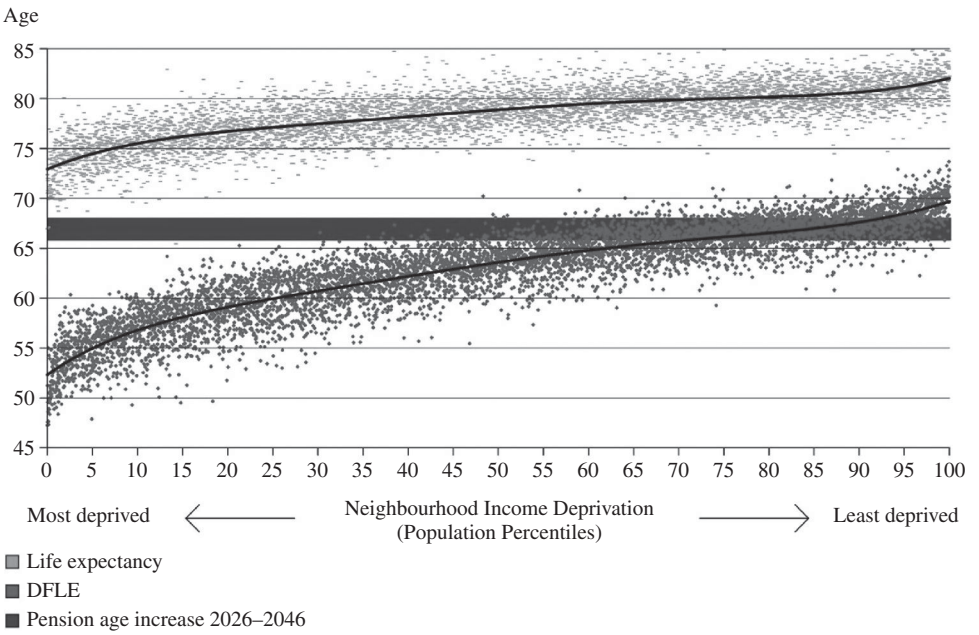


FIGURE 15.3: Life expectancy and disability-free life expectancy (DFLE) at birth, persons by neighborhood income level, England, 1999–2003.

April 2013. This new legislative framework currently applies to the Department of Health and the National Health Services organizations but, if these duties are upheld, may facilitate the prioritization of health equity throughout the health system.

While there has been explicit adoption of the review's approaches, efforts to further health equity have been undermined by some of the policy responses to economic crisis (11). We continue to work at the national level to influence and advise on social determinants approaches, and Michael Marmot's University College of London (UCL)-based Institute of Health Equity has provided many new analyses of approaches and evidence for government and other national organizations.

Locally, the impact has been even more significant. Seventy-five percent of local authorities have adopted strategies and approaches based on the English Review. Public health and other sectors across local governments are prioritizing health inequities and working together to implement innovative approaches on the SDH (12). The English Review prioritized the early years as the key area for intervention—as activity to reduce inequity here will have life-long benefit. The early years sector have in many cases taken up the challenge, and are recognizing the health impact of their activities and making collaborations across sectors (13).

Implementation work in England continues and, following an analysis on the potential role of health professionals in taking action on the social determinants, 22 health professional organizations have committed their workforces to greater activity within the healthcare sector (14).

THE EUROPEAN REVIEW

Following the publication of the English Review, the regional director for the European Region commissioned Michael Marmot and team to undertake a Review of the Social Determinants and the Health Divide across Europe. The review was set up to deepen understanding of the level of health inequity within the region, between and within countries, and to provide proposals for action based on the most up-to-date evidence from the region. The review also informed the strategies and approaches for the WHO Euro's Health 2020 program. As with other reviews, the European Review drew on global expertise and set up 13 task groups to inform the process. Based on the evidence provided by the task groups, the review grouped its recommendations into four themes: wider society, macrolevel context, systems, and lifecourse stages (Figure 15.4).

There is a clear emphasis on lifecourse stages. Evidence shows that inequity builds over the lifecourse, but in different ways at different stages of life. Interventions must be delivered in appropriate ways for different lifecourse stages and should aim to disrupt the accumulation of negative effects on health and well-being. The European Review has developed and extended the CSDH conceptual approach, and several new themes emerge from the review.

- Human rights are central to action on the SDH; human rights embody fundamental freedoms and the societal action necessary to secure them.
- In addition to addressing harmful influences, it is important to build on the resilience of individuals and communities; empowerment is central.

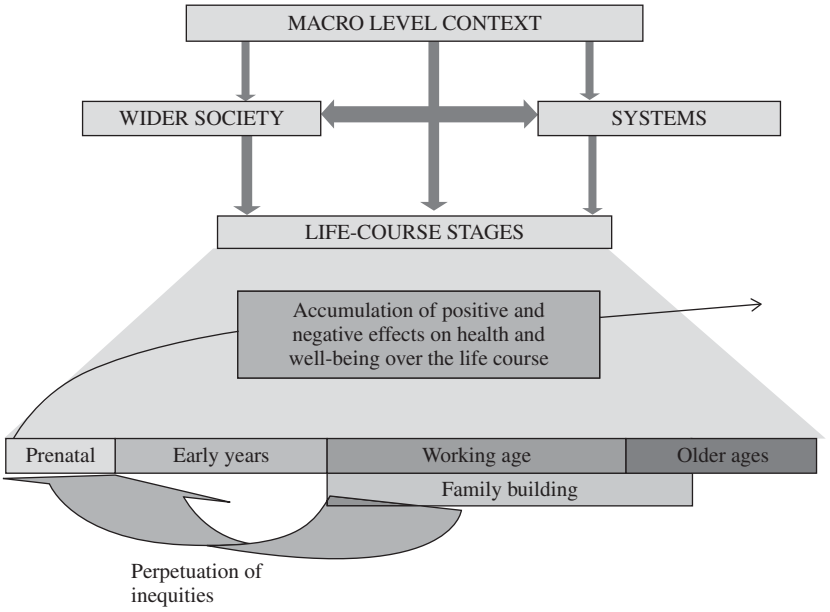


FIGURE 15.4: Broad themes of the review of social determinants and the health divide in the WHO European region.

- The lifecourse emerges as the right way to plan action on SDH; although the review emphasizes early childhood, action is needed at every stage of life; it makes strong recommendations for working and older ages.
- Protecting future generations from the perpetuation of social and economic inequities affecting previous generations is important.
- Intergenerational equity features strongly, in addition to intragenerational equity.
- Strong emphasis is needed on joint action on SDH, social cohesion, and sustainable development; all imply a strong commitment to social justice.
- Proportionate universalism should be used as a priority-setting strategy in taking action to address health inequity (3).

The review makes detailed recommendations for action in each of the themes outlined in the conceptual approach. There is clear evidence of the importance of social protection and transfers in reducing health inequities at the same time as many of these protection mechanisms are under threat from austerity. The ambition for the findings of the review are that the WHO, other international organizations, and national governments take forward the findings and approaches outlined in the review to shape strategies and actions and to foster the political prioritization and political leadership of health equity, which is so central to enabling reductions of health inequities.

While there has been much promising activity, there has also been a lack of progress in many arenas.

In Europe and the Americas, action on the SDH is better developed than in other regions. This may be an indication of a stronger political will, based on a longer history of social welfare and social

justice; more extensive evidence, based on causal relations between determinants and outcomes; and more developed data monitoring with surveillance and linked resources to take action (4).

In all regions, the global financial crisis has added urgency to consideration of dramatic, and often widening, financial inequities within and between countries. As standards of living decrease in many countries, and government revenues are tightened, we would argue that it is even more urgent that the distributional effects of all policies be taken into account in policy decision-making. However, this does not seem to have been a factor in many policy and economic responses to financial crisis. Recent analysis of evidence indicates that the economic crises since 2008 will likely worsen health and health inequity; adding to this, many austerity policies may further negatively impact health inequity, and the focus has been primarily on restoring economic growth, rather than prioritizing sustainable, equitable policies (11, 15).

BARRIERS AND CHALLENGES

In the previous few sections we have outlined the process and findings of three major reviews of the SDH. We have described how evidence has been translated into policy recommendations and is taken up by policymakers and political systems. We have described some of the impacts and potential further impacts of the findings of these reviews and suggested that there has been some clear evidence of success, but also continuing challenges to the social determinants approach and to efforts to tackle societal inequities. In this section we describe in more detail some of the challenges we have faced in translating evidence into policy and in advocating for the SDH approach.

THE NATURE OF EVIDENCE

Discussions with policymakers and those practitioners responsible for delivery indicate a clear need for evidence of interventions that have demonstrated success in tackling health inequity in a cost-effective way. Evidence of this type is scarce—evaluations of complex social interventions are rare; rarer still are those that can quantify cost and impact. It is still unusual for effective evaluations to be built into program design and budgets. The complexity of programs and outcomes makes assessment of impact difficult, and the necessary long-term nature of health impact means that health outcome data are particularly difficult to establish. The task is twofold—first we must persuade policymakers and practitioners that there is more than sufficient evidence of causal pathways; that we have robust evidence-based conceptual models that provide good confidence about effective action; that we know more than enough to take action in specific ways. Second, there must be greater focus on financing and planning good evaluations of interventions in ways that provide the types of evidence policymakers need. While the evidence on cost is hard to assess and demonstrate, where it does exist it can shift policies in unhelpful directions. We need to implement the most successful policies to reduce inequity, not to achieve efficiency or financial savings. The most desirable policy would do all three, but we must not prioritize efficiency over equity.

PRIORITIZING HEALTH EQUITY

Good leadership is critical to the success of action to tackle health inequity; without leadership there will be no effective prioritization of the issue, and action quickly slips off agendas. To support the development of this leadership, we in the academic/policy field must continue to make relevant, timely analyses; provide information of the scale and human and economic costs of health inequity; and provide relevant proposals for action that are aligned with wider government agendas. Political timescales are often short, and the arguments must be presented and represented with political cycles. Providing evidence of short-term gains is helpful in persuading politicians to invest in what is necessarily a long-term strategy with outcomes unlikely to be seen for many years.

Advocacy is important to achieving the necessary leadership—provision of evidence of need, indicators of trends, and wider analysis that can support policy leadership will all be critical to achieving prioritization, and academics, researchers, and advisors can take on a leading role.

PUBLIC SUPPORT

It is hard to ensure the necessary political focus on health inequity if politicians are unconvinced that the public are engaged in the debate. While the focus of debate remains on individual responsibility, the public see themselves as responsible for ill health and lower life expectancy, rather than understanding that there are also SDH, which influence health behaviors.

Public engagement in SDH is challenging; interest in inequity remains mostly focused on the healthcare system. Work that fosters public engagement and understanding of the scale and causes of health inequities would provide politicians with more motivation to prioritize the issue and would provide a greater scale of interest from other stakeholders who have responsibility for reducing health inequities—employers and local government, for instance.

MEASUREMENT AND MONITORING

Without measurement and monitoring of health distributions locally, nationally, and internationally it is unlikely that there will be anything like sufficient understanding or prioritization of health equity. Good systems of data collection, monitoring, and measurement of health and social determinants is critically important to activity on social determinants. In many regions and countries there are insufficient health data, and establishing data systems remains a clear priority for action. In countries such as England, which have had strong health equity measurement and monitoring systems, there is a need to protect the existing systems, which have been threatened as a result of cuts (16).

LOCAL, NATIONAL, AND INTERNATIONAL ACTIVITY

The social determinants evidence around the significance of wider societal processes has led some commentators to suggest that only major international and national policy levers are effective

in reducing health inequities. We suggest that action is necessary at every level and that local agencies and authorities can achieve a great deal, even without national political action; local activity in England has clearly demonstrated this.

EMPLOYERS AND PRIVATE SECTOR ENGAGEMENT

In the first edition of this book, Michael Marmot wrote in his chapter that it was studies of Whitehall that led him to make connections between levels of control at work—psychosocial factors and health outcomes. There is now a considerable and clear evidence base demonstrating the importance of good quality employment and good health. Poor quality work has a clear and measureable negative impact on health, and poor quality work is more frequently found in low-status jobs. While there is widespread recognition of the importance of good quality work and of employment generally, employers in both private and public sectors have been slow to make changes or become engaged with issues of health inequity more broadly. There is considerable scope for employers to take a greater role in efforts to reduce health inequity, and work to involve and engage them needs to be deepened.

CROSS-GOVERNMENT SUPPORT AND PARTNERSHIP WORKING

In previous sections we suggested that it has been difficult to achieve the necessary cross-sector approaches for effective action on social determinants. There have been some successful approaches and a great emphasis on mechanisms such as health in all policies to achieve this. However, it remains clear that cross-sector working requires exceptional leadership, motivation, and a shared sense of endeavor and collaborative partnerships. Health in all policy approaches need to be more than mechanistic assessments and should ideally have the support of the head of government.

ECONOMIC CONTEXT

Earlier in this chapter we briefly described evidence that showed that the economic crisis will likely have worsened health inequities, compounded in many cases by austerity policies and cuts to social protection programs. The focus of many international organizations and national governments seems solely on achieving economic growth, without consideration of the equity or the distributional impacts that may occur with economic growth. The narrow focus undermines endeavors to continue to focus on health equity. As the European Review has stated, "there is an even more pressing need for action on the SDH to ensure that a commitment to health equity survives and is enhanced."

CONCLUSIONS

It is still the case that most policy discussions of health are, in reality, discussions of healthcare. For policymakers, "health" is something that happens in the health sector, that is, healthcare. In Britain, now, and in several other countries, there is increasing recognition of the importance of SDH. This recognition builds on the movement to place Health in All Policies. The SDH approach requires health equity in all policies. All policies should be evaluated for their impact on the equitable distribution of health. The CSDH, the Marmot Review in England, and the European Review of Social Determinants and the Health Divide provide clear policy recommendations, which must be adapted to specific national or local contexts, and which, if followed, have the potential for significant positive impact on health equity. The hallmark of our approach, and perhaps one reason for take-up, is to base our recommendations, always, on the best evidence available. The political will required must be based on giving priority to achieving social justice.

REFERENCES

1. Marmot M. Multilevel approaches to understanding social determinants. In: Berkman L, Kawachi I, editors. Social epidemiology. New York: Oxford University Press; 2000. pp. 349–67.
2. Commission on Social Determinants of Health. Closing the gap in a generation: equity through action on the social determinants of health. Final report of the Commission on Social Determinants of Health. Geneva: World Health Organization, 2008.
3. UCL Institute of Health Equity. Review of social determinants and the health divide in the WHO Region: final report. Copenhagen, Denmark: WHO Regional Office for Europe, 2013.
4. Marmot M, Allen J, Bell R, Goldblatt P. Building of the global movement for health equity: from Santiago to Rio and beyond. Lancet. 2012;379(9811):181–8.
5. World Health Organization. Rio Political Declaration on Social Determinants of Health. 2011.
6. Leppo KO, E.; Pena, S.; Wismar, M.; Cook, S. Health in all policies: Seizing Opportunities, Implementing Policies: Ministry of Social Affairs and Health, Finland; 2013.
7. Chatterjee M. (SEWA). Personal communication.
8. The Marmot Review. Marmot Review Task Group Reports. 2010.
9. The Marmot Review. Fair society, healthy lives: a strategic review of health inequalities in England post-2010. 2010.
10. HM Government. Healthy lives, healthy people: our strategy for public health in England. London: Department of Health, TSO (The Stationery Office), 2010.
11. UCL Institute of Health Equity. The impact of the economic downturn and policy changes on health inequalities in London. London: Department of Epidemiology and Public Health, University College London, 2012.
12. UCL Institute of Health Equity. Available from: www.instituteofhealthequity.org.

13. UCL Institute of Health Equity. An equal start: improving outcomes in children's centres. London: Department of Epidemiology and Public Health, University College London, 2012.
14. UCL Institute of Health Equity. Working for health equity: the role of health professionals. London: Department of Epidemiology and Public Health, University College London, 2013.
15. Stuckler D, Basu S. The body economic: why austerity kills. London: Allen Lane; 2013.
16. Marmot M, Goldblatt P. Importance of monitoring health inequalities. BMJ. 2013;347:f6576.

INDEX