

Statistical Methods for Human Rights

Jana Asher · David Banks · Fritz J. Scheuren
Editors

Statistical Methods for Human Rights

 Springer

Editors

Jana Asher
1111 Jackson Ave
Takoma Park, MD 20912
jana@asher-resnick.us

David Banks
Institute of Statistics & Decision Sci.
Duke University, Durham, NC, USA
Durham 27708
banks@stat.duke.edu

Fritz J. Scheuren
NORC
University of Chicago
55 E. Monroe
Chicago II 60606
Scheuren-fritz@norc.uchicago.edu

ISBN: 978-0-387-72836-0

e-ISBN: 978-0-387-72837-7

Library of Congress Control Number: 2007931615

© 2008 Springer Science+Business Media, LLC

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper.

9 8 7 6 5 4 3 2 1

springer.com



Graciela Mellibovsky Saidler was a 29-year-old Argentine government economist. In 1976, she produced a statistical study on conditions in the slums of Buenos Aires which was so deeply embarrassing to the military dictatorship that it was publicly singled out by the Junta leader, General Jorge Videla, as an example of the infiltration of subversives into the government. Shortly afterwards, on September 25, 1976, she “disappeared.” In 1984, her father, Santiago, wrote a letter to the American Statistical Association, asking its help in determining her fate. In response, the ASA posted advertisements in Argentine newspapers offering a reward for information on her whereabouts. A few weeks later, the ASA received a letter from a former “death squad” member, Antonio Francisco Valdez, who claimed knowledge of her death. The ASA, in conjunction with the Committee on Scientific Freedom and Responsibility of the American Association for the Advancement of Science (AAAS), sent investigators to Buenos Aires to interview Valdez who, at the time, was imprisoned for

ordinary criminal charges. He gave a statement in which he confessed, in chilling graphic detail, to torturing and killing Graciela, referring to her as the “beautiful Jewess.” He also demanded an exorbitant sum of money to disclose the location of her grave. A few weeks later, he escaped and, after murdering his wife, was killed in a shoot-out with police.

More than three decades later, the ultimate fate of Graciela, like those of thousands of other Argentine desaparecidos, remains unknown. Her aging parents, Santiago and Matilde, have never given up their search for her.

It is fitting that this work, devoted to the use of statistics in the investigation of human rights violations, be dedicated to Graciela, who gave her life to bring truth through numbers.

Foreword

I was particularly pleased and honored when Jana Asher, David Banks, and Fritz Scheuren asked me to write some lines as a foreword to an uncommon compilation of papers dealing with all aspects of statistics in the service of Human Rights and Democracy.

No surprise if a large majority of contributors to this book are from the United States: as recalled by Tom Jabine and Doug Samuelson (see Chapter 9), as far back as 1982, the American Statistical Association transformed into a standing committee an Ad Hoc Committee on Scientific Freedom and Human Rights, created after Carlos Noriega and Graciela Mellibovski disappeared and, in all probability, were executed by the military regime in Argentina in the late 1970s. Of course, defending statisticians in peril is just one aspect of the work statisticians have to do to promote and develop an evidence-based assessment of human rights implementation.

As Tom and Doug also related in Chapter 9, it was not so easy to raise this concern at the level of the International Statistical Institute (ISI), which refused in 1985 to have a meeting on “Statistics, Statisticians and Human Rights” in the official scientific program of the ISI Centenary Session. At that time, it also refused to create a committee to review ISI’s policies on human rights. Contributed papers were, nevertheless, presented and an informal open meeting was organized.

Years after the events I have just recalled, when I served as the President of the ISI, its International Association for Official Statistics (IAOS) section organized, in Montreux (Switzerland) in September 2000, its biennial international conference on the topic “Statistics, Development and Human Rights.” More than 500 professional statisticians, governmental officials, policy analysts, and human rights practitioners attended. The Montreux conference was undoubtedly a major step in developing a broad approach to the issues at stake for the statistical community, and in Chapter 8 Jan Robert Suesser and Raul Suarez de Miguel describe a project that is a direct outcome of that conference.

This book—edited by Jana, David, and Fritz—is a welcome occasion to deal with many of the problems yet to be fully addressed. One of the most important of those problems is to demonstrate that it is possible to use statistical methods to measure human rights violations, included in the most extreme situations as genocides or large-scale conflicts (and moreover to contribute to update the definition of a genocide from a legal point of view) as explained in several chapters of the present book, for instance (a) Mary Gray and Sharon Marek in Chapter 2 devoted to genocides;

(b) Clyde Collins Snow et al. about the Guatemala civil war (1977–1986) in Chapter 5; (c) Romesh Silva and Patrick Ball in Chapter 6 devoted to the measurement of killings and disappearances in the long 15-year conflict in Timor-Leste.

Chapters 2, 5, and 6 are all applied illustrations of the methods presented by Herbert and Louise Spierer in their fundamental textbook *Data Analysis for Monitoring Human Rights* edited in 1993 by the American Association on the Advancement of Science (AAAS) and the network HURIDOCs (Human Rights Information and Documentation Systems). Building nicely on the work of Spierers are the conceptual models and other sophisticated statistical methods permitting in-depth analysis aiming at better understanding Human Rights violations as described in Chapter 4 by Jorge Luis Romeu.

Among the first users of such statistical applications are the judicial systems of many countries that now recognize the value of scientific evidence, including statistical analysis of data. In Chapter 10, Herbert Spierer and William Seltzer describe their experiences as experts in cases submitted to the International Criminal Tribunal for the Former Yugoslavia and the International Criminal Tribunal for Rwanda created by the United Nations Security Council; in their chapter, they recall that one of the first use of evidence based on quantitative data analysis was made during the Nuremberg Trial after World War II.

While these developments are dramatically needed to strengthen the values on which our societies rely on, it was important not to limit such a compilation on Statistics and Human Rights only to the problem of Human Rights violations. For years, international organizations have stressed the necessity to collect, analyze, and disseminate reliable figures in the field of human rights implementation taken in their broader definitions (including social, economic, and environmental rights), democratic processes, and governance improvement.

In the spirit of the declaration¹ of Amartya Sen, Dean of Trinity College, Cambridge, UK, and Nobel Prize Laureate (2000), there are many statistical compilations now of positive achievements in human rights. Among these are those of the UNDP (with the Human Development Reports produced yearly produced since 1991), the World Bank (with its programs on Public Sector Governance, Millennium Development Goals, or Anti-Corruption), and the OECD (with its projects on both Public Governance and Management, and Sustainable Development). Certainly

¹ “It’s impossible to reach a sustainable growth only by the means of an economic growth, but Human Development and respect of Human Rights and Democratic Rights, as well as the improvement of Governance are as well important!

Bold new approaches are needed to achieve universal realisation of human rights in the 21st century—adapted to the opportunities and realities of the era of globalisation, to its new global actors and to its new global rules. . . . Every country needs to strengthen its social arrangements for securing human freedoms—with norms, institutions, legal frameworks and an enabling economic environment. . . . Legislation alone is not sufficient. Poverty eradication is not only a development goal—it’s a central challenge for human rights in the 21st century. . . . Information and statistics are a powerful tool for creating a culture of accountability and for realising human rights. Activists, lawyers, statisticians and development specialists need to work together with communities. The goal: to generate information and evidence that can break down barriers of disbelief and mobilise changes in policy and behaviour.”

these organizations are active and important actors and they are to be commended for increasing the visibility of actions in that direction.

The Metagora project, which is presented in Chapter 8 by Raul Suarez de Miguel and Jan Robert Suesser, plays an original and interesting role in addition to all these other initiatives, mainly because of its bottom-up approach which completes and enriches the top-down approach that characterizes the work done by most international actors. The lecturers of this book will certainly be enthralled by the description of the pilot projects realized often in a difficult and sensitive environment. They will also be interested by the use of statistical methods for estimating casualties from wars presented by Beth Osborne Daponte in the Chapter 3 or the survey on Afghan refugee camps related by James Bell et al. in Chapter 7. Another application is proposed in Chapter 13 in the presentation by William Seltzer and Margo Anderson of the use—and misuses—of population data systems to target vulnerable subgroups and individuals; their discussion proposes important ethical issues and guidelines to governmental statistical agencies and the statistical profession more generally when they work within such systems.

Last but not least, the perspective offered by David Banks and Yasmin Said in Chapter 11 on New Issues in Human Rights Statistics is of great importance: they stress that a traditional conception of human rights which is focusing only on their civil and political aspect is not sufficient. For many actors in society, social and economical rights are fundamental. We find an echo to their concerns with the Millennium Development Indicators and Poverty Reduction Strategic Frameworks which are also tools at the service of human rights. In the Chapter 12, David J. Fitch et al. precisely present some statistical problems and remarks linked with the practical implementation of Millennium Development Goals.

I do not doubt that the lecturers of this book will find many incentives to think about the issues which are presented through all the valuable experiences compiled here and will be inspired by reading them, much as I was.

Jean-Louis Bodin
Inspecteur Général honoraire de l'Institut National
de la Statistique et des Etudes Economiques, France
Former President of the
International Statistical Institute (1999–2001)
Fellow of the American Statistical Association

Acknowledgments

The authors and editors of this book wish to thank Audrey Chapman, Stephen Hansen, Susan Hinkins, Art Kendall, Thomas Moore, Dean Resnick, Mark Schonlau, Françoise Seillier, and Yves Thibadeau for serving as referees (and in Yves' case, translator) for the chapters contained within. The editors further thank the Science and Human Rights Program of the American Association for the Advancement of Science and Duke University for institutional support of this project.

Contents

Part I Statistical Thinking on Human Rights Topics

- 1 Introduction** 3
Jana Asher
- 2 The Statistics of Genocide** 37
Mary W. Gray and Sharon Marek
- 3 Why Estimate Direct and Indirect Casualties from War?
The Rule of Proportionality and Casualty Estimates** 51
Beth O. Daponte
- 4 Statistical Thinking and Data Analysis:
Enhancing Human Rights Work** 65
Jorge L. Romeu

Part II Recent Projects

- 5 Hidden in Plain Sight: X.X. Burials and the Desaparecidos
in the Department of Guatemala, 1977–1986** 89
Clyde Collins Snow, Fredy Armando Peccerelli, José Samuel Susanávar,
Alan G. Robinson, and Jose Maria Najera Ochoa
- 6 The Demography of Conflict-Related Mortality in Timor-Leste
(1974–1999): Reflections on Empirical Quantitative Measurement
of Civilian Killings, Disappearances, and Famine-Related Deaths** . . . 117
Romesh Silva and Patrick Ball
- 7 Afghan Refugee Camp Surveys in Pakistan, 2002** 141
James Bell, David Nolle, Ruth Citrin and Fritz Scheuren

8 Metagora: An Experiment in the Measurement of Democratic Governance 157
 Jan Robert Suesser and R. Suarez de Miguel

Part III History and Future Possibilities

9 Human Rights of Statisticians and Statistics of Human Rights: Early History of the American Statistical Association’s Committee on Scientific Freedom and Human Rights 181
 Thomas B. Jabine and Douglas A. Samuelson

10 Obtaining Evidence for the International Criminal Court Using Data and Quantitative Analysis 195
 Herbert F. Spierer and William Seltzer

11 New Issues in Human Rights Statistics 227
 David L. Banks and Yasmin H. Said

12 Statistics and the Millennium Development Goals 241
 David J. Fitch, Paul Wassenich, Paul Fields, Fritz Scheuren, and Jana Asher

Part IV A Final Word of Warning

13 Using Population Data Systems to Target Vulnerable Population Subgroups and Individuals: Issues and Incidents 273
 William Seltzer and Margo Anderson

Author Index 329

Subject Index 333

Contributors

Margo Anderson, *Professor*

Margo Anderson is Professor of History and Urban Studies at the University of Wisconsin-Milwaukee. She teaches American history and quantitative historical analysis. Her research has focused on the history and political uses of population data systems in modern societies, particularly on census taking. She served as a member (1998–2003) and Chair (2000–2002) of the Council of the Interuniversity Consortium for Political and Social Research (ICPSR), and President of the Social Science History Association (2006).

margo@csd.uwm.edu

Jana Asher, *Statistical Consultant*

Jana Asher is a statistician that specializes in the collection and analysis of human rights violations data. Prior to becoming an independent consultant, she served as Senior Program Associate at the Science and Human Rights Program of the American Association for the Advancement of Science. Her work in human rights has included the design and implementation of a national survey on human rights abuses in Sierra Leone, and technical advice for truth commissions and tribunals for Perú, East Timor, Sierra Leone, and Kosovo.

jana@asher-resnick.us

Patrick Ball, *CTO and Director, Human Rights Programs*

Patrick Ball is Chief Technology Officer and Director of the Human Rights Data Analysis Group at the Benetech Initiative. He has spent 17 years designing databases and conducting quantitative analysis for truth commissions, non-governmental organizations, tribunals, and for United Nations missions in El Salvador, Ethiopia, Guatemala, Haiti, South Africa, Kosovo, Sierra Leone, Sri Lanka, and Perú.

pball@benetech.org

David L. Banks, *Professor of the Practice of Statistics*

David Banks obtained his Ph D in statistics in 1984 from Virginia Tech. Since then he has worked at several universities (Berkeley, Cambridge, Carnegie Mellon, Duke) and several federal agencies (National Institute of Standards and Technology, Bureau of Transportation Statistics, Food and Drug Administration). His interest in statistical aspects of human rights was sparked by I. R. Savage's presidential address to the American Statistical Association and has continued undiminished.

banks@stat.duke.edu

James Bell, *Russia, Ukraine, and Commonwealth Branch Chief*

James Bell is Division Chief for Europe–Eurasia in the United States Department of State’s Office of Research.

BellJ2@state.gov

Jean-Louis Bodin, *Inspecteur Général de l’INSEE honoraire (retired)*

Jean-Louis Bodin has worked for 40 years in the French national system of official statistics where he held various positions. From 1989 to his retirement in early 2006, he was involved in projects of technical assistance to transition and developing countries and in international activities and is still very active in that domain. He is an honorary member of the International Statistical Institute (ISI) and was President of the ISI (1999–2001). During his presidency, he was one of the promoters of the IAOS Independent Conference held in Montreux (Switzerland) on “Statistics, Development and Human Rights,” the first worldwide scientific conference on this topic. He is also a fellow of the American Statistical Association (ASA).

jean-louis.bodin@wanadoo.fr

Ruth Citrin, *Senior Analyst*

Ruth Citrin is a senior Near East analyst with the U.S. Department of State. Citrin previously worked with the Department of State Office of Research, where she managed public opinion polling in the Near East and Afghanistan. Citrin received her master’s and doctorate in international relations from The Fletcher School of Law and Diplomacy. She has worked and researched extensively in the Near East.

citrinr@state.gov

Beth Osborne Daponte, *Senior Research Scholar*

Beth Osborne Daponte is a senior research scholar at Yale University’s Institution for Social and Policy Studies. Her research interests include demographic estimation methods, evaluation approaches, human rights, and US food assistance and policy. In addition, she is currently studying law at Quinnipiac University.

beth.daponte@yale.edu

Paul J. Fields, *Associate Professor*

Paul Fields is on the faculty of Brigham Young University in the Department of Statistics. Previously he was on the faculty at the United States Naval Postgraduate School where he was Director of the International Health Resources Management Program (IHRM). As a nation-building program, the mission of IHRM was to increase the social, economic, and political stability of participating countries by creating the conditions that foster democracy. In the role of educator, consultant and diplomat, Fields worked with senior government officials to improve the health status of the populations in over 50 countries from Albania to Zimbabwe.

pjfields@stat.byu.edu

David J. Fitch

David J. Fitch currently lives in a little indigenous aldea in Guatemala where he moved out of concern for all the poor mothers of the world whose children are needlessly dying. His motivation comes from knowing his mother’s grief should he have

died. David reasons that experimentally evaluated programs in health and education will provide compelling evidence for a grant application to the US Millennium Challenge Corporation for the promotion of Millennium Development Goals. When a relevant report hit his computer in early 2005, he organized the session “Statistics and the Millennium Development Goals” held at the Joint Statistical Meetings in 2006. David hopes to make contact with others seeing a role for experimental evaluation of development programs.

dfitch@uvg.edu.gt

Mary W. Gray, *Professor*

Mary W. Gray is Professor of mathematics and statistics at American University, Washington, DC, and a lawyer. She is a long-time activist in Amnesty International and on the human rights committees of scientific societies. Mary is currently working on statistics and human rights projects in the Middle East and Latin America.

mgray@american.edu

Thomas Jabine, *Statistical Consultant*

Thomas Jabine is an independent consultant specializing in sampling, survey research methods, and statistical policy, with prior experience as a federal statistician at the U.S. Census Bureau, the Social Security Administration, and the Energy Information Administration. Several of his publications have dealt with the collection and analysis of statistical data relevant to human rights issues.

tjabine@starpower.net

Sharon Marek, *Graduate student, School of International Service, American University*

Sharon Marek is a graduate student in the School of International Service at American University. Her research interests include the use of Geographic Information Systems and statistical methods for human rights projects, and the impact emerging technologies have on global security arrangements.

smarek@american.edu

David Nolle, *Staff Specialist (Research Methodology)*

David B. Nolle is a staff specialist with the Office of Research in the U.S. Department of State, and he has been involved since 1974 with many public opinion survey research projects done around the world by the Office. His professional interests include public opinion research, ethnic relations, and research methodology.

dbnolle@att.net

José Mario Nájera Ochoa, *Chief, Forensic Medicine Department*

Nájera is Chief of the Forensic Medicine Department at the Guatemalan Attorney General's Office, in which he has worked for over a decade. He graduated from the Universidad de San Carlos de Guatemala. He holds a Diploma in Forensic Medicine, is a surgeon, and a law graduate.

naochojm@gmail.com

Fredy Armando Peccerelli, *Executive Director*

Fredy Peccerelli is a forensic anthropologist. He earned a master's degree from the University of Bournemouth, United Kingdom. He is Executive Director and a founding member of the Guatemalan Forensic Anthropology Foundation (FAFG). He is the current president and a founding member of the Latin American Forensic Anthropology Association (ALAF). He has worked in several investigations in Guatemala, Bosnia, and Croatia.

fredy.peccerelli@fafg.org

Alan G. Robinson, *Director, Forensic Anthropology Laboratory*

Alan Robinson is an archaeologist and forensic anthropologist. He earned a master's degree from University of Bradford, United Kingdom. He currently works in Guatemala as Director of the Forensic Anthropology Laboratory of the Guatemalan Forensic Anthropology Foundation (FAFG). He has worked in several investigations in Guatemala and Kosovo.

alan.robinson@fafg.org

Jorge Luis Romeu, *Research Professor*

Jorge Luis Romeu is Research Professor with the Department of Mechanical and Aerospace Engineering, Syracuse University and Director of the Juarez Lincoln Marti International Education Project. Romeu, born and raised in Cuba, has published two books on Cuban issues and has been a Fulbright Senior Speaker Specialist in Mexico, Dominican Republic and Ecuador. He served in the ASA and NACAE Human Rights Committees and has written and presented papers about HR subjects in AAAS, ASA, ORSA, NACAE, ASCE, and other professional association publications and meetings.

jlromeu@syr.edu

Yasmin H. Said, *Ruth L. Kirschstein National Fellow, Statistical Methodology Director, Computational Statistician*

Yasmin H. Said is Ruth L. Kirschstein National Research Fellow from the National Institutes of Health. She does alcohol modeling, agent-based simulation modeling, and major public policy work trying to minimize negative acute outcomes, including HIV/AIDS, related to alcohol consumption. Said is also Statistical Methodology Director of the Innovative Medical Institute, LLC and is a computational statistician by training. She is the editor of *Computing Science and Statistics*, is an associate editor of the journal *Computational Statistics and Data Analysis*, serves on the board of the Washington Statistical Society, and on the American Statistical Association Presidential Task Force on Science Policy. She is currently one of the editors of the six volume *Wiley Encyclopedia of Computational Statistics* and is writing a book *Controversies in Global Warming* and another *Statisticians of the Twentieth Century*. With colleagues she has developed testimonies on global warming for the House Committee on Energy and Commerce and to the House Subcommittee on Oversight and Investigations. She has also taught probability and statistics at The Johns Hopkins University in Baltimore, MD.

ysaid99@hotmail.com

Douglas A. Samuelson, *President, Operations Research Analyst*

Doug Samuelson is an operations research analyst and statistician. Since 1988, he has headed his own consulting company, InfoLogix, Inc. He was Chair of the American Statistical Association's Committee on Scientific Freedom and Human Rights, 1985–1988, and has participated in many human rights projects over the past 25 years. Recent projects include studies of mass egress and evacuations following disasters or attacks; uses and abuses of voting machines in the US; assessment of quality of simulation models of complex organizations; and organizational information processing and decision-making.

samuelsondoug@yahoo.com

Fritz Scheuren, *Vice President, Statistics*

Fritz Scheuren is a survey statistician, the 100th president of the American Statistical Association (ASA), and a senior consultant at NORC, an independent research arm of the University of Chicago. He has participated in many human rights projects over the last 11 years or so. Most of these are mentioned in his 2005 ASA Presidential talk, which appears in the December 2006 issue of the *Journal of the American Statistical Association*. Currently Fritz is leading a Millennium Challenge project for NORC in the Republic of Georgia.

scheuren@aol.com

William Seltzer, *Senior Research Scholar*

William Seltzer is Senior Research Scholar, Department of Sociology and Anthropology, Fordham University. In his long professional career, he has been active in demographic measurement, international statistics, professional ethics in both demography and statistics, studies in the misuse of population data systems, and in the use of quantitative data to document human rights abuses. After having retired as Director of the United Nation's Statistics Division in 1994 and joined Fordham, in addition to this research and writing, he has served as a consultant to the Office of the Prosecution, International Criminal Tribunal for Rwanda, a member of the National Research Council's Roundtable on Forced Migration, and Chair of the American Statistical Association's Committee on Professional Ethics.

seltzer@fordham.edu

Romesh Silva, *Statistician*

Romesh Silva is Statistician at the Benetech Initiative's Human Rights Data Analysis Group. He has led human rights data projects in Timor-Leste with the Commission for Reception, Truth and Reconciliation, in Sri Lanka with the Human Rights Accountability Coalition, in Northern India with Ensaaf, and in Chad and Bangladesh with Human Rights Watch. Romesh has also contributed to human rights data projects in Guatemala, Colombia, Burma, and Palestine.

romesh@benetech.org

Clyde Collins Snow

Clyde Snow is an independent forensic anthropology consultant who has helped in the investigation of human rights violations in over two-dozen countries throughout the world. He has served as an expert witness in national and international courts in the Hague, Argentina, and Ethiopia. Most recently, he has given expert testimony in Baghdad in the trial of Saddam Hussein for his genocidal “Anfal Campaign” against the Iraqi Kurds.

ClydeSnow@aol.com

Herbert F. Spirer, Professor Emeritus

Herbert F. Spirer is Professor Emeritus of Operations and Information Management at the University of Connecticut and former Adjunct Professor in the School of International and Public Affairs at Columbia University. He has been a consultant to the International Tribunal to the Former Yugoslavia for evidence concerning crimes against humanity. He is coauthor of *State Violence in Guatemala, 1960-1996, Making the Case*, and *Data Analysis for Monitoring Human Rights* (translated into three languages), and many other publications.

hspirer@optonline.net

Raul Suarez de Miguel, General Co-ordinator

Raul Suarez de Miguel is the manager of Metagora, an OECD-hosted international project aimed at enhancing evidence-based assessment of human rights and democratic governance. He organized the Montreux Conference on “Statistics, Development and Human Rights” held in 2000. He was head of the department of international affairs of the Swiss Federal Statistical Office, and spent 15 years in designing, organizing, and supporting cooperation projects in countries of Eastern Europe, the Balkans, and the South-Mediterranean area.

raul.suarez@oecd.org

Jan Robert Suesser, Statistician-Economist

Jan Robert Suesser, a statistician–economist, is working as a senior civil servant. He actively participated in the Montreux conference on “Statistics, Democracy and Human Rights” and in the Metagora pilot project focusing on methods, tools, and frameworks for measuring democracy, human rights, and governance. He is a member of the French major Human Rights NGO named “Ligue des Droits de l’Homme.”

jrswes@wanadoo.fr

José Samuel Suasnívar, Sub-executive Director

José Samuel Suasnívar is a forensic archeologist from the Universidad de San Carlos de Guatemala. He currently works as Sub-executive Director of the Guatemalan Forensic Anthropology Foundation (FAFG). He has worked in several investigations in Guatemala, Honduras, Bosnia, Croatia, and DRC. He is a founding member of the Latin American Forensic Anthropology Association (ALAF) and the author of over 40 articles.

jose.suasnavar@fafg.org

Paul Wassenich, *Ph.D. student*

Paul Wassenich is a development economist whose past work has focused on project impact evaluations at the World Bank and the Inter-American Development Bank. He is currently a Ph D student in the Department of Agricultural and Resource Economics at the University of California, Berkeley.

wassenich@are.berkeley.edu

Chapter 1

Introduction

Jana Asher

The world of international human rights law, practice, and policy is complex, and the role that statistical analysis plays within such a world is difficult to understand without some background knowledge as to what a human right is, how human rights law has evolved over time, and the different players that lead to the realization of that law. Additionally, the distinction between humanitarian and human rights emergencies is sometimes a subtle one, and humanitarian and human rights law and practice can overlap. This chapter introduces those concepts in order to enlighten the statistician or social scientist as to the role of quantitative measurement in human rights advocacy and enforcement.

This chapter begins in Section 1.1 with a brief definition of human rights. Section 1.2 gives an introduction to the history of the “modern” human rights movement—that is, the development of human rights law and protocol since World War II. Section 1.3 describes how nongovernmental organizations (NGOs) are placed in the human rights community. In Sections 1.4 and 1.5, the traditional role of statisticians in that community is highlighted and a history of human right violations data analysis projects is presented in two parts: first the history of the quantitative analysis of coded qualitative data is explored, and then the development of random sample survey methods within the developing world, in general, and for human rights violations data, in specific, is outlined. The chapter concludes in Section 1.6 with a discussion of the potential roles of statisticians in the human rights community of the future and a discussion of the structure of the remainder of this book.

1.1 Human Rights

The exact nature of human rights has been debated for millennia and is still a topic of debate today. There is, however, a current “working” definition of human rights that can be derived from existing international treaties and conventions. The creation of those treaties will be discussed in Section 1.2; here we note that existing international human rights treaties outline *civil and political rights* such as

- freedom of speech
- freedom of assembly

- freedom from slavery
- freedom from torture
- the right to a fair trial¹

and *economic, social, and cultural rights* such as

- equal wages for work of equal value without discrimination on the basis of gender
- time for rest and leisure and reasonable limitation on working hours
- the right to social security
- the right to free, compulsory elementary education
- the right to physical and mental health
- the right to participate in the cultural life of the community.²

Human rights abuses, then, are violations of international human rights law perpetrated by governments or organized regimes. In the case of civil and political rights, such abuses can include mass killings and disappearances, physical assault, torture, forced displacement, sexual assault, and unlawful detention. In the case of economic, social, and cultural rights, abuses are more difficult to pinpoint, but can include the seizing or destruction of property, forced/unpaid labor, the withholding of emergency food and/or medical relief, a culture of impunity for discrimination in the workplace on the basis of race, ethnicity, or gender, or corruption in a governmental entity which causes funding for education or social security to “disappear.”

In order for human rights to be universally attainable, international human rights conventions and treaties and national human rights law must be enforceable via a *policing body* and a *working court system*, so that human rights abuses potentially can be prevented and perpetrators of human rights abuses can be apprehended, tried, and punished. Although the international governmental system has been moving toward the creation of such a system for enforcing international human rights law, that process is far from complete as is shown in Section 1.2.

1.1.1 Human Rights and Humanitarian Law and Advocacy

A point of confusion is the relationship between human rights law and advocacy and humanitarian law and advocacy. International humanitarian law is specific to the treatment of soldiers and civilians during times of war. The main legal documents that define international humanitarian law are the Geneva Conventions,³ which outline fundamental rights of captured and/or injured combatants and civilians within

¹ See the International Bill of Human Rights for the text of the primary international human rights treaties related to civil and political rights (www.ohchr.org/english/law/index.htm, accessed May 3, 2007).

² See the International Bill of Human Rights for the text of the primary international human rights treaties related to economic, social, and cultural rights (www.ohchr.org/english/law/index.htm, accessed May 3, 2007).

³ See “Humanitarian Law” at www.ohchr.org/english/law/index.htm (accessed May 3, 2007).

war zones. Humanitarian advocacy most often takes the form of humanitarian aid or relief and can include the provision of food, water, medicine, shelter, and durable goods to victims of natural disasters (such as famines, earthquakes, tsunamis, and droughts) or man-made disasters (such as wars).

International human rights law is applicable both during times of peace and also during times of war, and concerns itself with the general behavior of governments instead of the specific behavior of military units and/or commanders. Human rights advocacy involves fact-finding on human rights abuses, advocacy for specific victims of abuse, and application of pressure on governments to cease abuse.

International humanitarian and human rights law may intersect—such as in the prosecution of government officials for war crimes. The same situation can represent both a humanitarian and a human rights crisis—for example, refugees fleeing a war zone—and therefore might require both humanitarian and human rights advocacy. Because this book is concerned with human rights specifically, humanitarian law and advocacy will be discussed only when pertinent to human rights law and advocacy.

1.2 A Brief History of Modern Human Rights Law and Practice

Mass human rights abuses were certainly not unheard of prior to World War II, even in the twentieth century. The Turkish genocide of Armenians during World War I (ICTJ 2003, Werful 1934), the Japanese Military's Nanking Massacre within China in December 1937—February 1938 (Yang 1999), and the Red Terror of Vladimir Lenin in Russia (Applebaum 2003) all illustrate that mass human rights abuses were a known phenomenon in the early 1900s. The international legal mechanisms in existence at the time, however, did not provide for either the prevention of mass atrocities or the prosecution of those who instigated them.

There were two clear reasons. First, the validity of *universal* human rights was not established. Legally, a right is a right because it is sanctioned and enforced by a state or inter-state organization—but an international treaty related to human rights had not yet been created.⁴ Second, the *sovereignty* of the state—or the state's right to self-governance and freedom from international intervention in its affairs (Reisman 1990)—was considered inviolate. Looked at from another angle, the belief was that any existing international law applied to states only, not individuals, and that individuals were a domestic concern (McDougal and Bebr 1964). That had long been a strong belief; even in the face of overwhelming human rights abuses perpetrated by a state against its own citizens, such as the examples given above, outside states traditionally had been reluctant to interfere. The concept of sovereignty, although still strong, has been challenged most in the last 15 years, as we will see in this chapter.

⁴ The Geneva Conventions are an example of Humanitarian Law, and thus technically are not human rights treaties.

The world's first attempt at an international organization of cooperation—the League of Nations—lasted only the span of the time period between the two world wars, failing, in the words of Das 1947, because “the nations and their leaders failed to maintain what was right and just, and their opportunism and national selfishness” won over cooperation (p. 53). As a result, it wasn't until after World War II—with the development of the United Nations—that human rights entered the international political arena in a meaningful way.

1.2.1 The United Nations

World War II saw a human rights crisis of unprecedented scale; approximately eleven million people had been killed by the Nazi party in its quest for a perfect “Aryan race.” Approximately six million of those people were Jews, and millions of additional soldiers were killed fighting that war (Green 1998). The Allied leaders, desiring to avoid a repeat of the bloodshed of that war, agreed that a new international organization that would serve as an arena for inter-state disputes and insure stability was needed in the post-war era. The efforts to create such an organization were spurred on by the US President Franklin Delano Roosevelt. During two meetings held in 1944 between the allied powers—well before the war had ended—Roosevelt led efforts to both begin what would become the charter of the United Nations and also establish the International Monetary Fund and the World Bank (Glendon 2001, p. 4). The allied partners met again in February of 1945, and their efforts bore fruit in the April–June 1945 San Francisco conference that would mark the beginning of the United Nations as an international body. Unfortunately, President Roosevelt died just one week prior to the beginning of that conference (Glendon 2001, p. 8).

Although Roosevelt's initial concept of the United Nations included a strong focus on human rights issues, this focus had faded in the pre-San Francisco meetings of the allied forces. Negotiations during and around the conference, however, led to the inclusion of human rights language in the UN Charter. First, US nongovernmental organizations exerted pressure on the US to make human rights a priority for the new organization, and the U.S. State Department, in response, championed the inclusion of a Human Rights Commission within the UN Charter. Second, many of the smaller countries attending the conference—including several Latin American countries—pushed strongly for human rights to be more prominently mentioned in the UN Charter. When the first pictures of concentration camps were released to the delegates at the convention on May 8, 1945, support for human rights goals within the United Nations grew. The end result was strong language in the preamble of the charter related to human rights:

to reaffirm our faith in fundamental human rights, in the dignity and worth of the human person, in the equal rights of men and women and of nations large and small... (Glendon 2001, p. 18)

Additionally, Chapter 1 of the charter pledges the member states to work within the United Nations to promote human rights and fundamental freedoms.

While the inclusion of human rights language in the UN Charter was a positive step for those seeking to ensure human rights, it was a shaky one at best. It was one thing to declare intent, and another thing altogether to act on that intent; the United Nations Charter has no provision for enforcement of its goals (Stettinius 1946). However, an almost immediate victory for human rights ensued in November of that year, 1945, when trials of the leaders of the Axis forces of World War II commenced in Nuremberg. While some may argue that these trials served only as an example of victor's justice, others point out the amazing achievement of the existence of these trials at all. As Bass (2000) points out,

If Nuremberg was not created out of perfect goodwill, it was still far better than anything else that has been done at the end of a major war. The pressures on the Allies to choose summary execution were enormous, but they resisted in the name of domestic liberal decencies. . . . In the end, America and Britain managed to produce something extraordinary. We have created nothing to compare with it since (p. 205).

At the end of 1945, worn out by two world wars, the world awaited the formation of the UN Commission on Human Rights, to see what would happen next.

1.2.2 Creation of the Universal Declaration of Human Rights and the Convention on the Prevention and Punishment of the Crime of Genocide

The period of solid progress, between the creation of the United Nations in 1945 through about 1954, was one in which human rights norms were defined and basic charters and treaties were drafted (Forsythe 1985). During that time period, two major human rights documents were created and accepted by the United Nations: the Universal Declaration of Human Rights and the Convention on the Prevention and Punishment of the Crime of Genocide.

The United Nations held its first General Assembly in January of 1946. In a bold move, the United States sent the late president's wife, Eleanor Roosevelt, as a delegate to that meeting. She proved herself an able spokesperson and diplomat, so much so that she was invited to be on the committee created by the United Nations to make recommendations as to the form and function of the Commission on Human Rights. At the Committee's meeting in April–May of 1946, she was elected its chair. While the committee made only two recommendations—that there should be an international bill of human rights and that the Commission members should be appointed by the UN on the basis of individual qualifications rather than state allegiance—it was successful enough that the first Commission on Human Rights met for the first time in January of 1947 (Glendon 2001, pp. 21–32). Again Eleanor Roosevelt was chosen to serve as its chair (Whiteman 1968). Out of that initial meeting sprang a sub-committee, led by Eleanor Roosevelt as well, that would draft the Universal Declaration of Human Rights. Through her able leadership and political skill, Eleanor Roosevelt successfully led the committee to the creation of a document that was adopted by the General Assembly of the United Nations on

December 10, 1948. The enormity of her achievement cannot be understated; in spite of the political tensions of the time, not one member of the United Nations voted against the declaration, and only eight members abstained (Nolde 1953).

During the same time as the development of the Universal Declaration of Human Rights, a different, separate treaty was being developed within the United Nations. Its inspiration stems from the experiences and intellectual work of one man—a Jewish lawyer from Poland named Raphael Lemkin (Keck and Sikkink 1998, p. 83). Lemkin was the scholar who created the word “genocide”—a combination of the Greek word for race, *genos*, and the Latin word for killing, *cide* (Gutman and Rieff 1999). Lemkin was convinced prior to World War II that the fate of the Armenians in Turkey during his boyhood would be repeated for the Jews. Sadly, while working on the staff of the chief prosecutor of the US at the Nuremberg trials, he learned that 49 of his family members had perished in horrific circumstances under the Nazi regime. After this, he lobbied heavily at the United Nations for a convention on genocide and contributed greatly to the concepts embodied in the final wording of the convention. On December 9, 1948—one day before the acceptance of the Universal Declaration of Human Rights—the United Nations approved the Convention on the Prevention and Punishment of the Crime of Genocide unanimously (Keck and Sikkink 1998, pp. 87–88).

Additional conventions and treaties were approved later, notably on issues such as prostitution (1949), refugees (1951), women (1953), stateless people (1953), and slavery (1953, 1956) (Forsythe 2000, p. 43). These must have been heady times for the members of the United Nations as basic rules were established and ratified. But as Kunz (1949) points out, “That human rights are enumerated in a constitution is no proof in itself that they exist (p. 320).” The very same countries that were ratifying conventions were breaking them within the sovereignty of their own territories. The period of enforcement of these norms had not yet come.

1.2.3 More Conventions and the Beginning of Action

The early to mid-1950s did not mark the end of the creation of conventions and human rights norms. Scholars and philosophers continued to debate the nature of human rights as the UN continued to refine its own definitions, through yet more conventions. Two of these conventions warrant discussion, as they can be considered the “culmination” of the UN’s “standard-setting” work: the Covenant on Civil and Political Rights and the Covenant on Economic, Social, and Cultural Rights, both approved by the General Assembly in 1966 (Donnelly 1981). These covenants served two purposes. First, while the Universal Declaration of Human Rights was extremely significant in the arena of international politics, it did not constitute international law, while the Covenants of 1966 did (Sohn 1979). Additionally, these two Covenants translate the principles given in the Universal Declaration into “precise legal language with detailed elaborations of exceptions, limitations, and restrictions” (Fischer 1977, p. 45). Second, the fact that rights were categorized and split into two different covenants—Civil and Political on the one hand, and Economic, Social,

and Cultural on the other—reflected a belief that the first type of rights only required negative action (e.g., to not torture citizens) and was more enforceable, while the second type of rights required positive action (e.g., to provide food for starving citizens). This distinction among human rights has not been universally accepted—for example, see Shue (1996)—but it is pervasive enough to be worth noting.

The 1966 Conventions were not ratified as quickly as the earlier human rights initiatives. As Robertson (1999) points out, it took 10 more years for enough states to ratify the conventions so that they could come into force. And between the acceptance of the Universal Declaration of Human Rights in 1948 and 1960, the United Nations Human Rights Commission had been forced to turn a blind eye on mass human rights abuses of states, instead focusing its efforts on paperwork. This lack of action on the part of the UN finally came to an end in the 1960s, when South Africa withdrew its membership. At last there was a state against which action could be taken with little to no political ramifications. By 1967, the General Assembly of the United Nations was calling for economic sanctions against South Africa, and the Human Rights Commission was ordered to study patterns of consistent human rights violations, as exemplified in South Africa (Robertson 1999 and Schwelb 1970). While the powers of the Human Rights Commission were limited to the study of state reports and the delivery of comments to the UN Security Council (Fischer 1982), this still represented a movement away from drafting of conventions and treaties to making those legal instruments actionable realities.

More evidence of the move toward action by the UN came in April and May 1968, with the creation of the Proclamation of Teheran at the International Conference on Human Rights held in Teheran (Schwelb 1970). The proclamation served as a reaffirmation by member states of their commitment to the full realization of not only political and civil rights, but economic and social rights as well, that is rights to food, water, health, and education (Donnelly 1981).

In the meantime, reports of gross violations of human rights outside of the South Africa example were not put on the agenda at the United Nations, or were addressed rather unevenly (Donnelly 1988). Why? First, as Bilder points out in his 1964 paper, at that time there were not “generally applicable and systematic international procedures or institutional machinery for actually receiving and investigating complaints of specific violations of human rights and taking appropriate steps to remedy them (p. 731).” Second, as discussed before, a strong undercurrent in international relations for millennia had been that of the sovereignty of the state and the principle of nonintervention. In fact, statements of state sovereignty had formed part of the UN Charter. As Donnelly (1986) argues, while the human rights norms established by covenants and treaties were international norms, the implementation of these norms was the sole responsibility of the state. The reason is that the principle of nonintervention in the state is a protective one; as pointed out in Slater and Nardin (1986), “the nonintervention principle sustains values such as national independence, diversity, and mutual restraint that are fundamental to international order (p. 86).” Even local leagues of nations, such as the Organization of American States, wrestled with the issue of nonintervention when facing the clear abuses by its members of their citizens (Ball 1961). Third, as pointed out by Henkin 1965,

Political and economic revolution, domestic instability, international tension, [the] Cold War—these have inhibited the growth of human rights in many parts of the world. Inevitably, these same forces shaped the United Nations and guided its politics and programs and its human rights activities no less than its other preoccupations (p. 510).

In other words, the world was in an unsteady state, and the political alliances and issues of the world carried into the workings of the United Nations, making it hard to achieve goals that went beyond noble words.

The forces at work fueling human rights abuses during this time were not just totalitarian governments—in fact, Mitchell and McCormick (1988) demonstrate that all governments of this time period committed at least some human rights abuses. But the worst state offenders were often aided and abetted by the foreign policy of the more democratic nations. For example, the international policy of the United States toward Central American countries throughout the 1970s and 1980s, including the provision of military aid to totalitarian regimes, led to mass human rights abuses, massacres, and potential genocide (see Danner 1994, Kaye 1997, Poe and Meernick 1995, and Porpora 1990). Indeed, the achievement or lack of human rights by the government of a developing state appears to have played little role in decisions by the U.S. State Department as to the allocation of aid to that state, and the reception of aid seems to have done little to improve the human rights record of that state (Carleton and Stohl 1987, McCormick and Mitchell 1988, Regan 1995).⁵ Forsythe (1990) goes so far as to label the period from 1953 to 1974 as one of “neglect” of human rights concerns by the US government, and the period from 1981 to 1988 (Reagan’s presidency) as one of “exceptionalism”, where the US felt that it stood as an example of human rights and that international standards were of no interest. Neier (1989) and Poe (1992) are not as pessimistic as Forsythe (1990) as to the intentions of the Reagan administration, supporting the claim that increased US foreign aid was linked to decreased evidence of human rights abuses and pointing out that toward the end of his presidency Reagan’s administration felt promoting international human rights was a major goal, albeit one that it applied inconsistently depending on its political agenda.

1.2.4 Truth Commissions, Tribunals, and Sovereignty

As the 1990s began, two trends—one positive and one positively frightening—converged to alter the landscape of human rights international law and practice. The first was a continuation of a small movement toward commissions created with the sole intent of determining the truth of what happened during a totalitarian regime. While the first of such commissions formed in 1971, in Uganda, and by the end of the 1980s, three such commissions had created reports, the crystallization of the concept of a “Truth and Reconciliation Commission” (TRC) did not occur until

⁵ There is an exception: according to Poe (1992), Jimmy Carter’s administration (1977–1981) made aid-allocation decisions based on human rights records of recipient states.

1991, with the formation of the Chile TRC. The concept of the TRC was further refined during the creation of the South African TRC in 1995, and the era of the modern TRC began (Hayner 2001, Gibsson and Gouws 1999).

Unfortunately, the proliferation of TRCs (over 20 at last count) reflects the more frightening trend apparent in the 1990s—that of massive levels of human rights violations in multiple countries. The international community was shocked to learn of camps much like the concentration camps of the Holocaust in existence in Bosnia in 1991. The situation in Yugoslavia over the next 9 years—including genocides, ethnic cleansings, massacres, and the forced migration of tens of thousands of people—led directly to the creation of the first international criminal tribunal since the Nuremberg trials, the International Criminal Tribunal for the Former Yugoslavia. Meanwhile, in Africa, trouble was brewing in Rwanda—trouble that culminated in the massacre of 800,000 Tutsi in April of 1994. Again, the United Nations responded by forming another criminal tribunal—in this case, the International Criminal Tribunal for Rwanda.⁶ In other countries, crimes were uncovered that were equally horrific, such as in the genocide of East Timor (Jardine 1999, Silva and Ball 2006, Chapter 6) and the massive and brutal human rights abuses in Sierra Leone (Conibere et al. 2004), leading to the formation of special courts by the UN and member states in both these countries to try and convict the worst of the perpetrators.

The call to action in the cases of Yugoslavia, Rwanda, and other countries owes its roots to sentiment that had been building in the international community over a number of years. McDougal et al. (1969) points out that “Deprivations of human rights visited upon one individual or group are increasingly perceived to be a personal deprivation for any observer and a potential threat to all freedom. Indeed, the knowledge is now pervasive that no people can really be secure in basic rights unless all peoples are secure (p. 237).” However, the initial United Nations responses in both Yugoslavia and Rwanda—the deployment of “peacekeepers” by the UN Security Council to prevent mass human rights abuses—were met with derision when those peacekeepers stood by while thousands of citizens were massacred in front of them. Nation states that had contributed soldiers to the peacekeeping forces were still not ready to forfeit their lives in the interference in another state’s issues (Barnett 1997). In contrast, the tribunals and special courts offered the real option of punishment for the offenders, even if the offenses were not stopped in the first place. The era of international policing of nation states’ human rights practices had begun, albeit shakily.

The United Nations’ movement toward action in the case of massive internal human rights was not limited to (failed) direct intervention in conflict and the creation of courts for trying the leaders of rogue states. In El Salvador and Guatemala, a UN delegation brokered human rights agreements with the two governments that included peace-building missions of UN staff as observers (Burgerman 2000). In other parts of the world, economic sanctions were used as a UN-sponsored form of international condemnation; Iraq after its invasion of Kuwait and mass slaughter of

⁶ See Bass (2000), Gourevitch (1998), Neuffer (2001), and Sudetic (1998) for more details on what happened in the former Yugoslavia and Rwanda.

Kurds serve as one example (Forsythe 1998, Mayall 1991). And in Kosovo in 1999, direct NATO bombing helped end Milošević control over Yugoslavia.

In spite of these advances for human rights abuse prevention and punishment, the situation was far from satisfactory. Not all human rights crises have been treated equally, as can be seen in the case of Chechnya (Cornell 1999) and China (Forsythe 1998). Important to note is that while mediation and individual cases of human rights abuses fell under the auspices of the Human Rights Committee and other similar committees of the UN, the UN Security Council controlled all peacekeeping or military actions taken by the UN (Alston 1983, Donnelly 1983). Therefore, international politics—most notably, in the structure of the UN Security Council—still determined which situations merited the peacekeeping attention of the UN, and which situations could be ignored. Such is the situation today, and such will be the situation as long as five countries on the Security Council—the United States, Russia, China, France, and England—have the individual power to veto any action of that council.

1.2.5 International Human Rights Law and Practice Today

Today, the United Nations' human rights machinery has grown to include multiple individuals, committees, and offices, coordinated by the Office of the High Commissioner for Human Rights (OHCHR). The UN Commission on Human Rights was replaced in 2006 by a UN Human Rights Council, which works with the support of the OHCHR. OHCHR also works closely with multiple United Nations programs and committees—including the United Nations Development Programme, the United Nations Children's Fund, the United Nations Educational, Scientific and Cultural Organization, the United Nations Population Fund, the United Nations Development Fund for Women, the Division for the Advancement of Women of the Secretariat, the Department of Peacekeeping Operations, the United Nations Staff College, and the Office of the United Nations High Commissioner for Refugees—to advance human rights in all UN activities. UN bodies that do not work directly with OHCHR but are concerned with human rights include the United Nations General Assembly, the Third Committee of the General Assembly, the Economic and Social Council, the International Court of Justice, and the International Criminal Tribunals for the Former Yugoslavia and Rwanda.⁷

Current UN policies on and protections of civil and political rights are moving more and more toward intervention and punishment. The trial of Slobodan Milošević at the Hague sent the warning out to potential human rights abusers that even their status as ex-leaders of countries would not protect them from international justice, and an international mechanism for punishing human rights offenders recently has been created via the International Criminal Court. As noted previously,

⁷ For more information on the complex network of bodies that comprise the United Nations' human rights system, please see www.ohchr.org (accessed May 3, 2007).

however, a matching international policing body does not yet exist, as the UN peacekeepers are deployed only at the request of the UN Security Council.

While some analysts consider the recent movement toward international justice a positive step, others still worry about the loss of sovereignty. As Semb (2000) points out, once the policy of nonintervention is softened, the border between “acceptable” interference and “political” maneuvering may become blurred, as, arguably, appears to have been the case in the US invasion of Iraq in 2003.

Regarding extradition, Dugard and Wyngaert (1998) point out that extradition to particular states might involve human rights abuses of the person extradited, especially in the case when the extradition is to a country that still practices the death penalty. As such, requests for the extradition of even the worst of the war criminals must be handled carefully (USCIRF 2005). And as Watkin (2004) points out, in the ever more complicated world reaction to terrorism, current humanitarian law may be outdated, and care must be taken to ensure that accountability for the deadly use of combat follows a norm based on human rights.

The UN policies and action on economic, social, and cultural rights have developed more slowly. In recent years, initiatives such as the Millennium Development Goals have begun to address the right to food, health, and education via development initiatives. The international human rights community at large has begun to approach development via a rights-based framework, and current projects and collaborations in that community—including multiple UN initiatives—are tackling the difficult task of developing indicators of such economic, social, and cultural rights (a task to which statisticians should endeavor to contribute).

In summary, when we consider the progress made in the understanding and realization of human rights—from the formation of the United Nations, to the creation of human rights covenants and treaties, to action against states that violate those treaties—we have come a long, long way in the span of a single lifetime. However, the abuses still occurring in the world today and the inequities experienced in access to food, education, employment, and health in developing countries show us how the next lifetime must be spent.

1.3 Nongovernmental Organizations

In our discussion so far about the history of human rights in the international arena, we have focused our attention on the actions of national and international governing bodies. This is not by accident; by definition, the worst of the possible human rights-related crimes—genocide, crimes of war, and crimes against humanity—can only be committed by governments, individuals acting under the authority of governments such as the Janjaweed in Darfur (Hagan and Palloni 2006, Human Rights Watch 2005), or organized regimes that may not have achieved recognized state status in the eyes of the international community, such as the Revolutionary United Front in Sierra Leone (Conibere et al. 2004). For this reason, an individual not associated with a government may be guilty of mass murder or racially motivated crimes, but she/he cannot be found guilty of genocide. Additionally, tribunals and

special courts typically limit their prosecution to those “high up” in the chain-of-command, rather than the individual soldiers or other governmental personnel that directly committed the crimes against citizens.

There is a category of organization, however, that plays an important role in the monitoring and promotion of human rights throughout the world, in spite of its lack of status as a governmental body. This type of organization is aptly known of as a nongovernmental organization, and can be national or international in nature. Because NGOs have played such a pivotal role in the creation of human rights treaties and the monitoring of human rights violations throughout the world, it is essential to understand how they operate in order to understand the international human rights community at all.

1.3.1 The International Committee of the Red Cross

The first and perhaps best-known international NGO is the International Committee of the Red Cross (ICRC). Although this NGO is not a human rights NGO per se—the ICRC concerns itself specifically with humanitarian law—it is presented here for completeness.

The ICRC owes its start to one man—a Swiss citizen named Henri Durant. In 1859, disgusted by the sight of wounded soldiers left to die on the battlefield of Solferino, he organized a neutral relief effort for the wounded. From his efforts sprang the progenitor of the International Committee of the Red Cross, an all-Swiss organization which, under the umbrella of the Swiss government, organized the first Geneva conference in 1864, where the first of the Geneva conventions was adopted. By 1875, Durant’s organization had officially adopted the ICRC name (Armstrong 1985).

Today the International Red Cross movement is comprised of three bodies: the ICRC, the national Red Cross societies, and the League of Red Cross Societies. The League of Red Cross Societies is the governing board for the national Red Cross societies. The ICRC branch of the International Red Cross is careful to keep its activities separate from the activities of the national branches. Indeed, the only connection between the ICRC and the national branches is that the national branches must be accepted as official by the ICRC in order to join the overall organization (Armstrong 1985, Forsythe 1976).

There are two good reasons for the separation between the League of Red Cross Societies and the ICRC. First, the national societies and the ICRC perform very different functions. While the national societies work in conjunction with their governments on health, nursing care, blood donations, relief operations, the ICRC works with man-made disasters as a neutral intermediary and is active in the formation of international humanitarian law (Armstrong 1985). Second, in the performance of their duties, the national societies are not fully private in the sense of being fully independent of governments. For example, the United States Red Cross has some leaders appointed by the government and is in charge of the nation’s blood supply under strict governmental regulation (Forsythe, 1976). If the ICRC is to

maintain its neutrality in international politics, it must remain separate from national governments.

The ICRC's relationship with the Swiss government, however, may call into question its true neutrality status. The ICRC is clearly an international NGO, but its majority funder has been the Swiss government, and until the 1970s the committee controlling the ICRC was strongly linked to the Swiss government (Armstrong, 1985). The International Red Cross conference, convened every 4 years, is in reality convened by the Swiss government in order that states may be invited and, at least in the past, the Geneva Conventions may be discussed, potentially modified, and approved. Because of the unique neutral standing of the Swiss government in the international arena, however, the ICRC's strong relationship with that state has not impeded its impact on the development and practice of humanitarian law (Forsythe 1976).

The impact of the ICRC on humanitarian world issues through the present day far outstrips that of any other international NGO. The ICRC has successfully legalized its already existing practice via the Geneva conventions at several points in history, and its role in conflict has expanded from the international battlefield, to civil unrest, to the monitoring of political prisoners. In short, it has successfully broken through the wall of sovereignty in a multitude of nations. One reason for this is the ICRC's strict stance on gaining the cooperation of the governments with which it works, and its willingness to release its reports to the offending government only, providing humanitarian assistance while not broadcasting what it finds to the rest of the international community. As such, cooperation with the ICRC is often desired even by governments responsible for the worst human rights and humanitarian offenses, as the assumption is that these governments are attempting to improve the situation via their ICRC relationship (Armstrong 1985, Forsythe 1976, Weissbrodt 1987).

The ICRC has come under criticism for this policy of cooperation with governments rather than condemnation of governments based on what it finds. But the result of its approach is an unparalleled access to those in greatest need of humanitarian assistance. It is the combination of the ICRC approach, and the more direct policing approach of the other international NGOs, that creates a reasonably effective system. As Armstrong (1985) states,

The ICRC has opted for a moral position that accepts one evil (an unjust government) in order to ameliorate another evil (some specific consequences for individuals of that government) . . . It is because other NGOs (and governments) are actively involved in trying to bring about the release of political prisoners and condemning the detaining governments that the ICRC is able to play its more restricted role without serious qualms of conscience. (pp. 640–641)

1.3.2 Amnesty International and the Age of the INGO

For many years the ICRC served as the only international NGO with clear and effective access to the international community of governments. This changed in 1961, when an English attorney, upset by the lack of legal defense for protesters in

Portugal under the Salazar regime, created a public campaign for amnesty for these political prisoners. It soon became clear that the need was greater than the initial focus, and Amnesty International (AI) was born. AI quickly became successful for two reasons. First, it would carefully pick a balance of cases from the “free” countries, the Communist block, and the developing countries to avoid accusations of political goals. Second, it would combine excellent research and documentation with letter campaigns directly to the offending government by its members (Scoble and Wiseberg 1974).

Because of its excellent reputation, AI was quickly granted “consultative status” by the United Nations, the Council of Europe, the Organization of African Unity, and the Organization of American States. This created a symbiotic relationship between these governmental organizations and AI; while the governmental organizations did not have the funds available to do the research into claims of human rights abuses themselves, they could rely on AI to provide the reports they needed to make decisions. In turn, AI could increase its effectiveness by gaining the direct ear of these governmental groups.

As time has passed, the number of human rights-oriented international NGOs has grown to over 200. Only a handful of these, however, have the reputation and means to draw the attention of national governments and international governmental bodies. Amnesty International and the ICRC continue to have impressive influence on the behavior of governments and have been joined by organizations such as Human Rights Watch, Physicians for Human Rights, and Freedom House (Forsythe, 2000). Typically, their influence is due to consultative status with governmental bodies.

Arguably, the work of international human rights NGOs could not occur without the complementary activities of over 10,000 national human rights NGOs. The national NGOs provide the backbone of effort for monitoring human rights throughout the world, as is described in the next section.

1.3.3 The National NGOs and the Role of NGOs in Human Rights Attainment

Thakur (1994) argues that the human rights NGO community plays three roles: norm generation, monitoring, and enforcement. In terms of norm generation, as pointed out in Sikkink (1998), NGOs were active members of the committees that drafted some of the most important international treaties on human rights issues, including the Universal Declaration of Human Rights. More recently, Physicians for Human Rights took an active role in the campaign and treaty to ban landmines. Often, the ideas for new norms stem from NGOs, even if the legalization of new norms must come from governmental bodies.

Monitoring and enforcement of human rights norms occur in tandem. Keck and Sikkink (1998) describe the role of international NGOs and national NGOs in monitoring and enforcement efforts in the context of transnational advocacy networks. In Keck and Sikkink’s model, local NGOs first become aware of issues and attempt

to resolve issues with their own local government. When these efforts to reach their government are blocked, they turn to their allies, the international NGOs, for assistance. Because the international NGOs have contacts outside of the government in question, they can both appeal to national governments in the “free” world, as well as international governmental organizations such as the UN, to do something about the situation. Those governments and governmental organizations, in turn, exert pressure upon the government in question to mend its ways. Perhaps economic sanctions are put in place, aid is withheld, or direct intervention is threatened. The end result, ideally, is a change in the government’s behavior. Because of the indirect way in which the local NGOs effect changes in their government—sending the message to the international community, which brings it back to the state in question—Keck and Sikkink call this the “boomerang pattern.”

Clearly NGOs—and especially national NGOs—have been essential for the monitoring and enforcement of human rights norms. As Johannsen (1983) points out,

In the absence of effective regional or global public institutions, nongovernmental organizations need to be utilized to hear claims in an impartial setting that transcends the vested interest of the national government involved, thus providing a quasi-public voice on human rights disputes. Such a voice partly overcomes the tendency to perceive claims as propaganda battles of one national government against another (p. 311).

In summary, there would be no modern movement toward international human rights without the efforts of nongovernmental organizations. In all aspects of human right achievement—norm generation, treaties and legal definitions, monitoring, and enforcement—NGOs have played pivotal roles and have allowed the progress seen in human rights justice in the world today. In the words of Thakur (1994),

The growing conviction that human rights do have a place in international relations, fed by NGOs such as Amnesty International, has been deeply corrosive of the inviolate rule of non-intervention . . . AI reports help to prepare the political atmosphere in the UN for resolutions authorizing . . . intervention (p. 155).

1.4 Quantitative Projects Based on Qualitative Data

Within the complex world of state governments, international governmental organizations, such as the Organization of American States and the United Nations, international NGOs, and local NGOs lies what Donnelly (1989) terms the “global human rights regime (p. 206)” and what Keck and Sikkink (1998) call the “global advocacy network.” Each member of the regime forms a point in an interconnected web without which the system would collapse. There is one more type of player in this global network that, to this point, we have not discussed: the conscience-driven scientist or scientific organization.

While the work of NGOs is unparalleled in its rooting out and reporting of human rights abuses, it is the technology developed by social scientists, physical scientists, and mathematicians that arm the international human rights community with better and better tools for proving their assertions and making their cases valid in the eyes

of the world. While some of these scientists work within particular organizations, most notably the United Nations, others provide services to a multitude of members of the network, serving as a strengthening material that helps hold it together. Within recent history, forensic anthropology has helped explain massacres in Argentina and the Philippines (Corillon 1989), as well as in Iraq and the former Yugoslavia. Geneticists have aided the reunification of children of the disappeared with their biological families in Argentina as well (Corillon 1989). Psychologists and psychiatrists have worked to assist survivors of abuses overcome their experiences, and medical personnel have fit prosthetic devices to children in Sierra Leone that have had limbs amputated.

The role of statisticians in this community has varied from analyzing data on totalitarian regimes for use by truth and reconciliation commissions or as evidence at legal trials to creating survey methodology within the developing countries for the measurement of good governance, attainment of human rights, and/or abuses of those rights. Of particular interest in this document is the role statisticians, social scientists, and computer scientists have played in the collection and analysis of human rights violations data. Two basic methodologies have evolved for the collection and analysis of those data: (1) the coding of existing qualitative data on human rights abuses and their subsequent analysis via multiple systems estimation and (2) random sample surveys. In this section, we will review the conversion of qualitative data on human rights abuses into quantitative data, and their subsequent analyses.

The history⁸ of the analysis of qualitative human rights violations data begins with the local NGOs that had collected stories of abuses within their own countries. Traditionally, as described above, these stories would then be passed on to international NGOs, which would make them available to governmental organizations such as the United Nations, who would then exert pressure on that country's government to improve the situation. At the onset of the Truth and Reconciliation Commissions, such qualitative data was deemed one of the more useful of the truth-seeking tools available, and many TRCs sponsored data collection efforts of their own.

The qualitative data obtained in this process came, for the most part, in the format of verbatim narratives by the victims of violations or the witnesses to violations. Because they were stories recorded verbatim, the abuses recorded would be described in whatever wording and language the storyteller chose. In order to turn such data into quantitative data, two important problems would have to be overcome. First, a decision as to what information to keep and what to throw away would have to be made, as a finite number of fields would be recorded for each event mentioned in the narrative (e.g., the name of the narrator, the name of the victim, the gender of the victim, the type of abuse experienced, etc.). Second, a "controlled vocabulary" for each field, especially the "type of abuse" field, would have to be created in order to effectively convert a potentially infinite number of separate abuse descriptions into a finite number of abuse codes (e.g., assault, torture, property destruction, etc.). Once these problems were overcome, a process would need to be created by which

⁸ This history was compiled using information from an interview with Patrick Ball on March 29, 2005.

the qualitative stories were converted into quantitative data; this process would be heavily expert-dependent as each story was subject to interpretation. Over time, this process was labeled “coding” by the human rights community. The quantified data would need to be entered into a database and “cleaned” to remove duplication of events, perpetrators, or victims. Finally, the data would be ready for analysis. Developing this methodology was no small task.

The effort toward turning human rights qualitative data into quantitative data suitable for statistical analysis first began as part of the Human Rights Information and Documentation Systems, International (HURIDOCS) project. Initiated in 1979 and formally established in 1982, the HURIDOCS project was begun by a small consortium of NGOs that saw promise in the burgeoning world of computers for a universal information handling system for human rights data. Over the years the HURIDOCS network has created multiple tools for information storage, including a standardized terminology and format for recording human rights information in databases (www.huridocs.org/brocheng.htm, March 30, 2005).

The HURIDOCS “events standard formats”, a controlled vocabulary for human rights events, stemmed from a 1985 survey of 40 human rights organizations, performed by HURIDOCS, as to their needs for information processing and storage. The survey determined that a universal systematic method for documenting human rights abuses, one that would allow information to be shared across organizations, was one of the main needs at that time. A HURIDOCS task force was convened in order to create such a system, and by 1993 it had published *Huridocs Standard Formats: A Tool for Documenting Human Rights Violations*. Shortly after its publication, an information specialist from Chile, Ricardo Cifuentes, created a form of database software to match the recommendations in the HURIDOCS manual. In 1994, a meeting jointly organized by HURIDOCS and the American Association for the Advancement of Science’s Science and Human Rights Program (AAAS-SHRP) was convened to discuss the design of databases for human rights violations data, and the HURIDOCS formats were revised in 1998 (www.huridocs.org/esfbackg.htm, March 30, 2005). As of 2007, HURIDOCS remains an active network.

The HURIDOCS effort is notable in that it was the first attempt toward creating a list of fields to be recorded and a controlled vocabulary for those fields. The HURIDOCS team, however, was not able to consider quality control issues in the coding or data entry process as part of their initial work. The beginning of quality control standards stemmed from another source.

During the time that the initial HURIDOCS task force was creating their standard formats, a separate initiative was taking place at the Los Angeles offices of a Latin American NGO named El Rescate. El Rescate had been providing assistance to refugees fleeing the civil war in El Salvador, and in the early 1990s led a multi-organizational effort to create a database of human rights violations that occurred during the El Salvador civil war. Their database was the first to link human rights violations directly to the responsible perpetrators and was relied on by the United Nations Mission Observers in El Salvador in assessing what had happened, as well as the El Salvador Truth Commission. The El Rescate database is, in fact, still used by the U.S. Department of Justice today

to determine the validity of requests for asylum on the grounds of political persecution in El Salvador (see www.hartford-hwp.com/archives/47/062.html, www.elrescate.org/humanrights.php?cat=elsalvadorwar#accountability, March 30, 2005). El Rescate's work is also notable because they were the first human rights database project to develop measures of inter-rater reliability for their coding process. In other words, they were the first to determine the probability, for a given abuse, that two separate coders would assign it the same abuse code. The subsequent database work of many NGOs and Truth and Reconciliation commissions was influenced by the El Rescate team.

A third, separate human rights database effort of the early 1990s was beginning on the other side of the world, in Bosnia, where there was a desire to create a registry of war crimes victims who were killed during the presidency of Slobodan Milošević. While this was initially an independent project, it became an office under the Bosnian government in 1995, and broke off into an NGO in April of 2004. Most notable about this project is the sheer size of the database created. Since 1992, the organization has collected one million pages of typed documents, from which it has generated 300,000 names of war victims and 30,000 names of perpetrators (www.idc.org.ba/basicprinciples.html, March 30, 2005).

Much of the development of these methodologies, however, can be credited to the efforts of Dr Patrick Ball, who worked within several organizations over the past 15 years. His first work was in El Salvador for the Salvadoran Human Rights Commission. Influenced by the work of El Rescate, he created a database entry project for data that had already been collected by multiple NGOs (Ball et al. 2000). During later work for the Haitian National Commission for Truth and Justice, semi-structured interview data were coded and then entered into a database under Ball's direction. Within this project, advances were made in the data collection phase by Margaret Weagers on the design of the form on which information was collected and in the training of the interviewers. Within the context of the collection of qualitative data, she was able to introduce methods that improved the professionalism and quality of the project.

The next big advance was in the data analysis phase, which until that point had been constrained to the type of exploratory data analysis outlined in Spierer and Spierer (1993). The advance occurred while Ball was working on behalf of AAAS-SHRP in Guatemala for the International Center for Human Rights Research. The challenge in Guatemala was to combine three separate sets of qualitative data into one database, accounting for duplication of events and overlap of actors (perpetrators and victims). With advice from Fritz Scheuren and Bill Seltzer, Ball used a combination of Bishop et al. (1975) and Marks et al. (1974) to develop a triple-system estimation procedure for the Guatemala database (Ball et al. 1999). Through that work, Ball was able to support the assertion that the violence in Guatemala amounted to genocide of the local indigenous populations (Ball 2000).⁹

⁹ Around the same time, Alan Zaslavsky was recommending the use of dual systems estimation techniques to a researcher attempting GIS mapping of rights violations. There is no evidence, however, that the analysis was to be applied to qualitative data processed via the methods described in this chapter (email from Alan Zasklavsky, March 1, 2004).

Ball finished his work with the Guatemala for the International Center for Human Rights Research in 1999. At that point in history, AAAS-SHRP was advising other Truth Commissions on appropriate database methodologies. Ball was also beginning work on the situation in Kosovo, traveling to Albania with Fritz Scheuren in 1999 to collect data on refugees. During this time, he also began to work with the American Bar Association Central and East European Law Initiative, and in 2000 the American Association for the Advancement of Science co-published a report that included a multiple systems estimation analysis much like the one Ball did in Guatemala (AAAS-SHRP and ABA-CEELI 2000). The next year, Ball and I redesigned those models, using a more involved log-linear modeling procedure. The result was a document that formed the basis for Ball's testimony at the trial of Slobodon Milošević in the Hague (Ball et al. 2002).

Since that point, several researchers have contributed expertise to improving the overall methodology for going from qualitative statements to data analysis. Ball and I collaborated again on a report for the Truth and Reconciliation Commission of Peru in 2003, developing methodology that was again based on log-linear models for multiple systems estimation (see Ball et al. 2003). Romesh Silva, working in East Timor and Sri Lanka, has contributed greatly to the understanding of inter-rater reliability models and controlled vocabularies. Ball has gone on to create the Human Rights Data Analysis Group (www.hrdag.org), initially housed at AAAS-SHRP and currently housed at the Benetech Initiative (www.benetech.org). Work there has focused on improvements in record linkage, inter-rater reliability, and databases and database interfaces.

1.5 Random Sample Survey Methods

NGOs and other collectors of human rights violations data have taken longer to embrace the direct collection of quantitative data via random sample surveys than the use of coded qualitative data for creation of human rights violations statistics. This is perhaps due to the need to maintain qualitative data collection; the inherent emotional punch of a human rights violations narrative makes advocacy efforts more effective than the use of statistics only. There is also the issue of background knowledge of the techniques of random sampling and the ability to collect the data in the manner the methodology demands. In spite of those challenges, random sample methods have become an invaluable tool for human rights advocacy over the past several years.

In order to understand some of the issues underlying the random sample survey of human rights violations, we begin with the development of random sample survey methods in general, and then discuss issues related to random sample surveys within developing countries, as human rights surveys can be considered a subcategory of those efforts.

1.5.1 History

Random sampling is a relatively new technique in data collection. The first generation of “modern” statistical sampling began at the turn of the century (Kish 1995) with the publication of Kiaer’s (1895) paper, “The Representative Method of Statistical Surveys.” Within that time period, two competing methods for achieving representativeness within a sample were used. The first, purposive selection, involves the deliberate selection of a sample for which particular characteristics of the underlying population—e.g., age, gender, race, and ethnicity—are present in correct proportions. The second, random sampling, involves the random selection of population members.

Both forms of sampling were accepted before Neyman’s (1934) paper, “On the Two Different Aspects of the Representative Method,” showed that purposive sampling did not equate to representative sampling, whereas a large enough random sample, as proven by theory, is “representative” of the underlying population in terms of all population characteristics. Random sampling started to become more common in the 1940s, when the U.S. Census Bureau began to employ this technique. Around the same time, Mahalonobis (1944) ran the first large-scale random sample survey in the developing world, in India on the farming of jute.¹⁰

During the 1940s and 1950s, as understanding of the power of random sampling grew, survey methodologists began to understand that survey data from the developing world, with the exception of India, were of questionable to low quality (Brass 1996). Statistical offices in developing countries (again with the exception of India), however, were slow to turn to random sampling techniques, due to lack of sample frames and infrastructure (Casley and Lury 1981 (p. 9)).

Since that time, however, great strides have been made in improving the collection of data in the developing world. Notably, the United Nations, through its Sub-Commission on Statistical Sampling (1947–1951) and Specialized Agencies did a great deal to spread an understanding of survey sampling during the 1950s and 1960s (Casley and Lury 1981 (p. 9), Murthy 1978). Between the 1960s and 1980s, the United Nations continued to support random sample survey work through funding international and national surveys (Casley and Lury 1981 (pp. 12–13)). Even today, the United Nations continues this tradition, most recently with the development of a manual for performing household surveys in developing countries (United Nations 2005).

Another notable effort toward improvement of data collection in the developing world is the University of Michigan Institute for Social Research Sampling Program for Survey Statisticians. Initially founded and run by Leslie Kish, that program is dedicated to training statisticians from developing countries in probability methods for developing complex random samples and has trained more than 500

¹⁰ The first documented use of multiple systems estimation for human populations also occurred in India; see Sekar and Deming (1949).

participants from over 100 countries since its inception.¹¹ Additional training efforts have included International Statistical Institute (ISI) workshops at the ISI biennial meetings, which are often targeted toward statisticians from developing countries and for which funding for statisticians from developing countries is provided when possible.

The available literature on surveys in the developing world reflects the growth of not only survey sampling methodology, but also questionnaire design methodology. While outlining advances in those topics, the literature gives examples of surveys on agricultural yield (Carlson et al. 1980, David 1988), voting behavior (Birmingham and Jahoda 1995), fertility and mortality (Adlakha et al. 1980, Madigan et al. 1976, Sabagh and Scott 1967), health (Bergsten 1980), nutritional status of children (Som 1997), living standards measurement (Chander et al. 1980), peasant attitudes (Frey 1963), population estimation (Lauriat 1967), induced abortion rates (Tezcan and Omran 1981), child labor (Ashagrie 2000), and contraception (Goldman et al. 1989). Of all of these surveys, several major international efforts stand out, including the World Fertility Survey (Casley and Lury 1981 (p. 12), Huntington et al. 1993, Little 1988, Vaessen 1978, Verma et al. 1980), the various survey-related projects of the World Bank,¹² and the Metagora human rights diagnosis project (Metagora, 2005). Through this literature, we can explore the issues and potential solutions associated with random sample data collection in the developing world.

1.5.2 Data Collection Issues

The developing world has benefited greatly from the use of random sampling techniques. The reason, as is pointed out by several authors (Adlakha et al. 1980, Bergsten 1980, Casley and Lury 1981 (p. 5)), is that census data and registry data (such as birth or death rates) available in developed countries are infeasible to gather or are simply not gathered in the developing world. Such data are needed, however, for governmental and UN planning, humanitarian assistance, and determination of guilt for tribunals and special courts. As pointed out in Mitchell 1965a,

Survey research techniques are generally accepted in most countries as the only means by which systematic and reliable data can be collected and analyzed on a wide variety of social and economic topics (p. 7).

Through efforts undertaken throughout the last half-century, random sample data collection problems in the developing world have been explored and better understood. While initially social scientist and survey methodologists tried to transfer survey methods as-is from the developed world to the developing world,

¹¹ For more information on this program, see www.isr.umich.edu/src/si/fellowship.html (accessed May 5, 2007).

¹² For example, the Living Standards Measurement Study and the Africa Household Survey Database (www.worldbank.org/LSMS/ and www4.worldbank.org/afr/poverty/databank/default.cfm respectively, accessed May 5, 2007).

they soon discovered that this was not an optimal strategy. As Bulmer and Warwick (1983) points out, while the obstacles to data collection present in the developing world are not that different from the obstacles found in the developed world, the extent of those problems is quite different. From Frey (1970),

Not only are the problems more severe, but their existence is more conspicuous. One can ignore a mouse-like problem, but when it assumes elephantine proportions, one ignores it only at the peril of being overwhelmed (p. 184).

Several notable issues in data collection in the developing world are apparent throughout the literature available. Some are listed here:

- Developing countries have high birth and mortality rates and high mobility in some cases. Any demographic data collected can quickly become obsolete as a result (Casley and Lury 1981 (p. 1)).
- Developing countries may lack the survey infrastructure that is found in developed countries: field workers, data processing facilities, and survey methodologists. The cost of surveys in developing countries may be unexpectedly high due to this lack of infrastructure (Yansaneh and Eltinge 2000).
- Due to difficult environmental conditions, political upheaval, and lack of knowledge of appropriate sampling techniques, developing countries either lack sampling frames completely or have sampling frames with significant errors (Bergsten 1980, Carlson et al. 1980, Mitchell 1965b). These errors can take the form of missing units, duplications, and inaccuracies/errors in recording (Casley and Lury 1981 (p. 53)).
- Due to environmental or political conditions in a developing country, or lack of sampling frames, interviewers may encounter a great deal of difficulty in traveling to and finding households (Bergsten 1980, Birmingham and Jahoda 1995, Carlson et al. 1980, Casley and Luch 1981 (p. 2), Murthy 1978).
- In some survey projects in the developing world, there is little to no attempt at pre-testing the questionnaire to be used (Carlson et al. 1980, Mitchell 1965b). In some data-gathering projects, there are no questions on the data collection form, leaving the structure of the questions up to the individual interviewer in the field (Carlson et al. 1980). In short, proper questionnaire design techniques are not utilized (Casley and Lury 1981 (pp. 74–91)).
- In some surveys in the developing world, answers to survey questions are taken verbatim without the use of source codes, leading to more error and greater interviewing time (Carlson et al. 1980).
- Quality checking of survey responses in the field may be nonexistent in many developing countries (Carlson et al. 1980).
- Commonly, developing countries contain multiple ethnic groups that speak a diverse set of languages. Often questionnaires are developed without consideration of how particular questions will be interpreted or understood across cultural groups, or whether particular concepts are directly translatable at all (Carlson et al. 1980, Jabine 1996, Mitchell 1965b). Language translation is not performed as part of pre-testing the questionnaire, but is left to the interviewer in the field, resulting in problems of equivalence of questions across languages (Bergsten 1980, Mitchell 1965b). Additionally, cultural sensitivities may differ from group

to group, and will not be addressed with sensitive question techniques due to on-the-fly translation (Mitchell 1965b).

- Developing countries have a larger percentage of illiterate or semi-literate respondents to surveys. Typically the interviewers, however, are fairly or very well educated in comparison to the rest of the population. This can cause issues in question formation (Birmingham and Jahoda 1995, Carlson et al. 1980), and shyness due to the “superior” social standing of the interviewer (Mitchell 1965b).
- In spite of the relatively high education levels of interviewers compared to respondents, the interviewers available are often not qualified to perform pre-testing or are untrained in interviewing techniques. Personnel may have only a high school education (Bergsten 1980, Carlson et al. 1980, Murthy 1978).
- In many developing countries, it may be close to impossible to insure privacy for the respondent to a survey; this can be especially problematic for sensitive questions (Bergsten 1980, Birmingham and Jahoda 1995, Mitchell 1965b).
- There is often underreporting of infant and childhood death in retrospective surveys in the developing world: the general consensus is that this underreporting is the result of reference period error, age misreporting, respondent forgetfulness, and/or concealment of events (Adlakha et al. 1980).
- The desire to please the interviewer may be quite high in comparison to Western cultures, leading to response bias in the answers if there is no pre-testing of questions or the interviewers are not trained carefully (Mitchell 1965b).
- The processing and analysis of the data may take too long for the collected data to be useful for the purpose for which it was collected (Casley and Lury 1981 (p. 14)).

The list of pertinent issues to data collection efforts in the developing world may seem overwhelming to the reader. Fortunately, there are some advantages to survey work in the developing world compared to survey work in the developed world as well:

- If appropriate steps are taken to gain permission of traditional authorities to interview members of their villages, nonresponse can be extremely low (Bergsten 1980, Casley and Luch 1981 (p. 125)).
- If the available sampling frame is lacking in a direct enumeration of households, but contains a reasonably accurate list of villages, then villages serve as a natural sampling unit in the sample design. Enumerating a random sample of villages on the spot is typically not difficult, yielding an excellent household survey (Bergsten 1980).

While the list of advantages is rather short, there are potential solutions to most, if not all, of the problems encountered during random sample surveys in the developing world. The literature is rich with advice.

With regard to sample design, one of the warnings given in Casley and Lury (1981) is that a survey methodologist in the developing world should “avoid indiscriminate transfer of technology” (p. 79). Carlson et al. (1980), Fitch et al. (2000), and Scheuren (2002) also warn the survey methodologist that data collection in the developing world is different; a sampling technique used in the developed

world may be infeasible or difficult to implement in the developing world. Casley and Lury issue the warning to keep the sample design simple (1981, pp. 13–14). Other sources talk of modifications to sample design used in developed countries to aid data collection, such as the EPI method of sampling (Bennett et al. 1994). Finally, Hyman et al. (1967) suggest the use of expert informants in regions in lieu of a complicated sample of individuals.

Other warnings and suggestions in the literature are related to questionnaire design. Because of the existing complications in the developing world, multi-use surveys are more impractical than in the developed world (Casley and Luch (1981, p. 44)). And because of variations in cultures within a developing country, “verbatim” questionnaire design techniques, where the interviewer is only allowed to read exactly what is on the questionnaire, should be avoided (Casley and Luch (1981, pp. 85–86)). Jabine (1996) points out that cognitive research techniques used for pretesting questionnaires can be used in any country, developed or not. Casley and Luch (1981) make the strong argument that field testing of the questionnaire must be done, and that translations of the questionnaire must be tested for equivalence of meaning before the survey begins (p. 97). Finally, several authors have tried randomized response for sensitive questions on surveys in developing countries with some success (Madigan et al. 1976, Tezcan and Omran 1981).

The issue of recall within a time period and date recall is especially problematic in developing countries, but Carlson et al. (1980) point out that agricultural cycles and timing are well understood across different cultures in developing countries. Casley and Luch (1981) suggest limiting the size of the recall period (pp. 79–81), while Goldman et al. (1989) report success using physical calendars to collect time information. Adlakha et al. (1980) discuss the issue of obtaining time or reference period specific data in detail. They suggest the use of multi-round surveys for fertility rates, and questions about pregnancy and birth history over a number of children for determining child morbidity and mortality.

Several authors emphasize the importance of adequate training of local personnel (Carlson et al. 1980, Kish 1996), but warn that those personnel with better training are likely to seek higher wages and “greener pastures” in the developed world. Care must be taken to train those committed to using their expertise within their own countries (Kish 1996, Verma 1996). In the context of training, if the trainer speaks a different language than the staff to be trained, excellent translation is required to allow the training to be effective, and written materials must be available in all the languages to be used (Kish 1996, Verma 1996). Such training may be needed to convince local people to use techniques such as random sampling at all (Jabine 1996).

Finally, Forster and Snow (1995) discuss the use of computer-based questionnaires in developing countries, asserting that the advantages in error control outweigh the costs.

Each of these suggestions has positive and negative features. Survey methodologists within developing countries often have little financial support in their efforts and must choose cost-effective techniques to implement their surveys at all. Exactly which suggestions and techniques will be most valuable depend deeply on the type of data to be collected and the country in which the collection will occur.

1.5.3 Analysis

A discussion of random-sample surveys in developing countries would not be complete without acknowledging advances in the analysis of data from such surveys. As Brass (1996) points out,

the great increase in the number of sample surveys, particularly during the last 30 years, has been an important factor in generating new ideas for analysis (p. 452).

The literature provides information on two types of advances in data analysis specific to random data from developing countries.¹³ The first is to collect data at a detailed level but analyze the same data at an aggregate level. The example given is within an agricultural survey, where the desired information is income level of farmers, but the data collected is on crops planted and expenses of crop production. Those data were much easier for the interviewers to collect and the farmers to report on than a total income, which would require bookkeeping that many farmers did not do (Carlson et al. 1980).

The second advance is labeled “indirect estimation” and discussed in Adlakha et al. (1980) and Brass (1996). Basically, indirect estimation involves the imputation of one type of information (e.g., fertility rates) from data gathered on a separate but related topic (e.g., age and sex profiles of the population). The authors advocate specifically for the use of survey and/or census data on sex, age, and mortality to determine number of births per woman; the use of survey reports on live children and maternal age to determine maternity histories; and a combination of census, survey, and/or registry data to determine mortality rates. In this way, the authors overcome the issues with collected reference period data from individuals in developing countries.

1.5.4 Random Sample Surveys on Human Rights

The history of random sample surveys specifically for human rights violations is relatively brief compared to surveys in developing countries in general. Although some of those surveys have been performed by governmental organizations (e.g., see U.S. State Department 2004), there have also been random sample surveys performed by nongovernmental organizations and private research groups. The ICRC has been involved in several such surveys (for examples see Greenberg Research Inc. 1999a, b). Women’s Rights International has done some excellent work developing surveys about sexual assault (Jennings and Swiss 2000, 2001), Roberts et al. (2004) and Burnham et al. (2006) explore the mortality of Iraqi citizens after the US invasion in 2003, and Physicians for Human Rights in particular has sponsored many

¹³ The author of this chapter only found information on two analysis techniques specific to developing countries. Clearly a variety of analysis techniques not mentioned here are applicable to data from surveys within developed and developing countries alike, and information on such techniques is widely available.

random sample survey projects (see Amowitz et al. 2004, Physicians for Human Rights 2001).

Benetech recently completed a survey of mortality for use by the Commission for Reception, Truth and Reconciliation in East Timor (see Silva and Ball (2006) and Chapter 6), and the American Bar Association sponsored a national survey of human rights abuses in Sierra Leone (Mejia 2006). Survey projects in Mexico, the Philippines, and South Africa completed under the Metagora Project (Metagora (2005), Chapter 8) are particularly noteworthy in that two of them—the survey in the Philippines and the one in South Africa—focused on the right to land, a decidedly economic right.

In spite of all of the work described above, technical expertise and methodology tailored to random sample surveys on human rights violations is in its infancy. Part of the reason for this is that often human rights violations are occurring in war zones; the feasibility of a random-sample survey in the middle of a war is not high (Jennings and Swiss 2000). Additionally, ethical issues related to human rights data collection are particularly thorny. Special care must be taken to ensure that the benefits of the survey to the target population are not outweighed by the potential harm of the survey to that population. Human rights surveys typically involve the reporting of traumatic information, and the potential for re-traumatizing the respondent in the process of the interview is significant if appropriate precautions are not taken. Such precautions include the use of questionnaire design methods for eliciting sensitive information, proper training of interview personnel, and potentially the use of counselors in the field.

Another ethical issue is the potential harm that could come to respondents if their participation in the survey is revealed to the wrong authorities. In some cases, simply having an interviewer present at a respondent's house increases the likelihood of harm coming to that respondent. In other cases, if proper methods for ensuring confidentiality are not used, data might be accidentally released to the wrong people.

In spite of the inherent challenges, however, the science of human rights surveys continues to improve, as Chapters 6 and 8 of this book can attest.

1.6 The Current Volume

The reader has now been grounded in human rights law and practice and has also explored some history of the statistical work that has taken place within the human rights community. The remainder of this book provides a snapshot of both the work being done by statisticians and statistically oriented professionals to promote and defend human rights today, and also some of the roles statisticians can play in the promotion and protection of human rights in the future.

In this first section of the book, we explore human rights topics from a statistician's point of view. Mary Gray and Sharon Marek debate how quantitative information may or may not assist us in defining genocidal events. Beth Osborne DaPonte discusses the measurement of excess civilian mortality

during conflicts—that is, mortality above what one would expect normally for the population—and the responsibility of governments for limiting the impact of war upon civilian life. Finally, Jorge Romeu describes statistical techniques for debunking the theory that human rights violations are a necessary but unfortunate consequence of rapid social development.

The next section of the book outlines three country-specific projects in human rights measurement, and one international project that explores the measurement of human rights, democracy, and good governance. Clyde Collins Snow leads an analysis team through data on the “disappeared” in Guatemala. Romesh Silva and Patrick Ball use a combination of random sample survey techniques and multiple systems estimation techniques to estimate civilian deaths during the Indonesian occupation of East Timor. James Bell and his coauthors present a survey of Afghan refugees describing their living conditions in Pakistan. Finally, Jan Robert Suesser and Raul Suarez introduce us to the Metagora project and demonstrate how that project proves measurement of governance is possible.

The third section of the book explores both the history of the intersection of human rights and statistics and also possibilities for the future. Tom Jabine and Doug Samuelson introduce us to the origins of the American Statistical Association Committee on Scientific Freedom and Human Rights. Herb Spierer and Bill Seltzer discuss the development of testimony for the newly created International Criminal Court (ICC), and David Banks and Yasmin H. Said present the possibility of donors using cost-benefit analyses to prioritize social development areas. A consortium of authors introduce the potential for improved monitoring and evaluation of development programs designed to meet the Millennium Development Goals—that chapter is based on an invited session on the same topic presented during the Joint Statistical Meetings in 2006. Inasmuch as development can be seen in a human rights framework—as the fulfillment of the right to employment, health, education, food, water, and shelter—this chapter might represent the strongest contribution that statisticians can make toward the promotion of human rights in the future.

Last but not least, lest the reader should forget that statistics can be used for good purposes or for evil ones, in the final section of the book, Bill Seltzer and Margo Anderson explore how population estimates have been used in the past to further genocidal plans.

This book is intended to be accessible to a wide audience of social scientists, lawyers, epidemiologists, demographers, statisticians, and other interested professionals. If the reader, after completing this book, is convinced of the fact that statisticians have multiple essential roles to play in ensuring the development of an equitable and just world, then the goal of the editors and authors of this book has been achieved.

Acknowledgment The author thanks David Banks, Audrey Chapman, Stephen E. Fienberg, and Fritz Scheuren for their comments on earlier versions of this chapter, Marcia Asher for her review of the final version of this chapter, and her son Duncan Groff for his patience and support during the writing and editing of this chapter.

References

- Adlakha AL, Sullivan JM, Abernathy JR (1980) Recent trends in the methodology of demographic surveys in developing countries. *Proceedings of the section on survey research methods of the American Statistical Association*, pp 60–65
- Alston P (1983) The alleged demise of political human rights at the UN: A reply to donnelly. *International Organization* 37:537–546
- American Association for the Advancement of Science Science and Human Rights Program (AAAS-SHRP), American Bar Association Central and East European Law Initiative (ABA-CEELI) (2000) Political killings in Kosova/Kosovo, March–June 1999: a cooperative report by the Central and East European law initiative of the American Bar Association and the science and human rights program of the American association for the advancement of science. ABA Central and East European Law Initiative, Washington DC
- Amowitz L, Kim G, Reis C, Asher J, Iacopino V (2004) Letter from Iraq: a population-based assessment of health and human rights in southern Iraq. *Journal of the American Medical Association* 291:1471–1479
- Applebaum A (2003) *Gulag: a history*. Doubleday, New York
- Ashagrie K (2000) Current progress in implementing new methods and conducting innovative surveys for measuring exploitation of children. *International Association for Official Statistics (IAOS) Conference “Statistics, Development and Human Rights”* (www.portal-stat.admin.ch/iaos2000/ashagrie_final_final_paper.doc accessed March 2, 2005).
- Armstrong JD (1985) The International Committee of the Red Cross and political prisoners. *International Organization* 39:615–642
- Ball MM (1961) Issue for the Americas: Non-intervention v. human rights and the preservation of democratic institutions. *International Organization* 15:21–37
- Ball P (2000) The Guatemalan Commission for historical clarification: intersample analysis. Chapter 11 in Ball P, Spierer H, Spierer L (eds) *Making the case: investigating large scale human rights violations using information systems and data analysis*. American Association for the Advancement of Science (AAAS), Washington
- Ball P, Asher J, Sulmont D, Manrique, D (2003) How many Peruvians have died? An estimate of the total number of victims killed or disappeared in the armed internal conflict between 1980 and 2000. AAAS, Washington
- Ball P, Betts W, Scheuren F, Dudukovich J, Asher J (2002) Killings and refugee flow in Kosovo March–June 1999: a report to the international criminal tribunal for the former Yugoslavia. AAAS, Washington
- Ball P, Spierer HF, Spierer L, (eds) (2000) *Making the case: investigating large scale human rights violations using information systems and data analysis*. AAAS, Washington
- Ball P, Scheuren F, Seltzer W, Spierer HF (1999) Multiple or N-System estimates of the number of political killings in Guatemala. *Proceedings of the section on government statistics and section on social statistics*. American Statistical Association, Alexandria, pp 156–160
- Barnett MN (1997) The UN Security Council, indifference, and genocide in Rwanda. *Cultural Anthropology* 12:551–578
- Bass GJ (2000) *Stay the hand of vengeance: the politics of war crimes tribunals*. Princeton University Press, Princeton
- Bennett S, Radalowicz A, Vella V, Tomkins A (1994) A computer simulation of household sampling schemes for health surveys in developing countries. *International Journal of Epidemiology* 23:128–1291
- Bergsten JW (1980) Some sample survey design problems in Syria, Nepal and Somalia. *Proceedings of the section on survey research methods of the American Statistical Association*, pp 72–74
- Bilder RB (1964) The international promotion of human rights: a current assessment. *American Journal of International Law* 58:728–734
- Birmingham WB, Jahoda G (1955) A pre-election survey in a semi-literate society. *Public Opinion Quarterly* 19:140–152

- Bishop YMM, Fienberg SE, Holland PH (1975) *Discrete multivariate analysis: theory and practice*. MIT Press, Cambridge
- Brass W (1996) Demographic data analysis in less developed countries: 1946–1996. *Population Studies* 50:451–467
- Bulmer M, Warwick DP, (eds) (1993) *Social research on developing countries: surveys and censuses in the third world*. University College London Press, London
- Burgerman SD (2000) Building the peace by mandating reform: United Nations- mediated human rights agreements in El Salvador and Guatemala. *Latin American Perspectives* 27:63–87
- Burnham G, Lafta R, Doocy S, Roberts L (2006) Mortality after the 2003 invasion of Iraq: a cross-sectional cluster sample survey. *The Lancet* 368:1421–1428
- Carleton D, Stohl M (1987) The role of human rights in U.S. foreign assistance policy: a critique and reappraisal. *American Journal of Political Science* 31:1002–1018
- Carlson BA, Kindel KK, Turner AG (1980) Survey design for control of measurement errors in developing countries. *Proceedings of the section on survey research methods of the American Statistical Association*, pp 66–71
- Casley DJ, Lury DA (1981) *Data collection in developing countries*. Oxford University Press, London
- Chander R, Grootaert C, Pyatt G (1980) Living standards surveys in developing countries. *Proceedings of the section on survey research methods of the American Statistical Association*, pp 83–88
- Conibere R, Asher J, Cibelli K, Dudukovich J, Kaplan R, Ball P (2004) Statistical appendix to the report of the Truth and Reconciliation Commission of Sierra Leone: a report by the benetech human rights data analysis group and the American Bar Association Central European and Eurasian law initiative to the Truth and Reconciliation Commission. (<http://hrdag.org/resources/publications/SL-TRC-statistics-chapter-final.pdf> accessed May 4, 2007)
- Corillon C (1989) The role of science and scientists in human rights. *The Annals of the American Academy of Political and Social Science* 506:129–140
- Cornell SE (1999) International reactions to massive human rights violations: the case of Chechnya. *Europe–Asia Studies* 51:85–100
- Danner M (1994) *The massacre at El Mozote*. Vintage Books, New York
- Das T (1947) Human rights and the United Nations. *The Annals of the American Academy of Political and Social Science* 252:53–62
- David IP (1988) Sampling strategy for agricultural censuses and surveys in developing countries. *Agricultural Statistics* 2000, pp 83–95
- Donnelly J (1981) Recent trends in UN Human Rights activity: description and polemic. *International Organization* 35:633–655
- Donnelly J (1983) The human rights priorities of the UN: a rejoinder to Alston. *International Organization* 37:547–550
- Donnelly J (1986) International human rights: a regime analysis. *International Organization* 40:599–642
- Donnelly J (1988) Human rights at the United Nations 1955–85: the question of bias. *International Studies Quarterly* 32:275–303
- Donnelly J (1989) *Universal human rights in theory and practice*. Cornell University Press, Ithica
- Dugard J, Van den Wyngaert C (1998) Reconciling extradition with human rights. *American Journal of International Law* 92:187–212
- Fischer DD (1977) The international protection of human rights. *Proceedings of the Academy of Political Science* 32:44–55
- Fischer DD (1982) Reporting under the Covenant on Civil and Political Rights: the first five years of the Human Rights Committee. *American Journal of International Law* 76:142–153
- Fitch DJ, Molina R, Barrientos T (2000) Weighting of survey data: is the view that such should not be attempted in developing countries reasonable? *Proceedings of the section on survey research methods of the American Statistical Association*, pp 764–769
- Forster D, Snow RW (1995) An assessment of the use of hand-held computers during demographic surveys in developing countries. *Survey Methodology* 21:179–184

- Forsythe DP (1976) The Red Cross as transnational movement: conserving and changing the nation-state system. *International Organization* 30:607–630
- Forsythe DP (1985) The United Nations and human rights, 1945–1985. *Political Science Quarterly* 100:249–269
- Forsythe DP (1990) Human rights in U.S. foreign policy: retrospect and prospect. *Political Science Quarterly* 105:435–454
- Forsythe DP (1998) Human rights fifty years after the universal declaration. *PS: Political Science and Politics* 31:505–511
- Forsythe DP (2000) *Human rights in international relations*. Cambridge University Press, Cambridge
- Frey FW (1963) Surveying peasant attitudes in Turkey. *Public Opinion Quarterly* 27:335–355.
- Frey FW (1970) Cross-cultural survey research in political science. *The methodology of comparative research*. Free Press, New York, pp 173–294
- Gibsson JL, Gouws A (1999) Truth and Reconciliation in South Africa: attributions of blame and the struggle over apartheid. *American Political Science Review* 93:501–517
- Glendon MA (2001) *A world made new: Eleanor Roosevelt and the Universal Declaration of Human Rights*. Random House, New York
- Goldman N, Moreno L, Westoff CF (1989) Collection of survey data on contraception: an evaluation of an experiment in Peru. *Studies in Family Planning* 20:147–157
- Gourevitch P (1998) *We wish to inform you that tomorrow we will be killed with our families: stories from Rwanda*. Farrar, Straus and Giroux, New York
- Green R (1998) Non-combatant deaths in WWII (www.holocaust-history.org/~rjg/deaths.shtml accessed March 18, 2006).
- Greenberg Research Inc. (1999a) Country report: Bosnia-Herzegovina. ICRC worldwide consultation on the rules of war. International Committee of the Red Cross, Geneva
- Greenberg Research Inc. (1999b) Country report: South Africa. ICRC worldwide consultation on the rules of war. International Committee of the Red Cross, Geneva
- Gutman R, Rieff D (eds) (1999) *Crimes of war: what the public should know*. W. W. Norton and Company, New York
- Hagan J, Palloni A (2006) Death in Darfur. *Science* 313:1578–1579
- Hayner PB (2001) *Unspeakable truths: confronting state terror and atrocity*. Routledge, New York
- Henkin L (1965) The United Nations and Human Rights. *International Organization* 19:504–517
- Human Rights Watch (2005) *Entrenching impunity: government responsibility for international crimes in Darfur*. Human Rights Watch, New York
- Huntington D, Mensch B, Toubia N (1993) A new approach to eliciting information about induced abortion. *Studies in Family Planning* 24:120–124
- Hyman HH, Levine GN, Wright CR (1967) Studying expert informants by survey methods: a Cross-national inquiry. *Public Opinion Quarterly* 31:9–26
- International Center for Transitional Justice (ICTJ) (2003) *The applicability of the United Nations convention on the prevention and punishment of the crime of genocide to events which occurred during the early twentieth century: legal analysis prepared for the international center for transitional justice* (www.ictj.org/images/content/0/9/092.pdf accessed April 30, 2007)
- Jabine TB (1996) Comment on “Developing Samplers for Developing Countries.” *International Statistical Review* 64:153–156
- Jardine M (1999) *East Timor: Genocide in paradise*, 2nd edition. LPC Group, Chicago
- Jennings PJ, Swiss S (2000) Statistical information on violence against women during the civil war in Liberia (www.portal-stat.admin.ch/iaos2000/jennings_final_paper.doc accessed Feburary 25, 2005)
- Jennings PJ, Swiss S (2001) Supporting local efforts to document human rights violations in armed conflict. *The Lancet* 357:302–303
- Johannsen RC (1983) Human rights in the 1980s: revolutionary growth or unanticipated erosion? *World Politics* 35:286–314
- Kaye ME (1997) The role of Truth Commissions in the search for justice, reconciliation and democratisation: The Salvadorean and Honduran cases. *Journal of Latin American Studies* 29:693–716

- Keck ME, Sikkink K (1998) *Activists beyond borders: advocacy networks in international politics*. Cornell University Press, Ithica
- Kiaer AN (1895) *The representative method of sample surveys*. English translation 1976, Statistik Centralbyro, Oslo
- Kish L (1995) The hundred years' war of survey sampling. *Statistics in Transition* 2:813–830
- Kish L (1996) Developing samplers for developing countries. *International Statistical Review* 64:143–152
- Kunz JL (1949) The United Nations Declaration of Human Rights. *American Journal of International Law* 43:316–323
- Lauriat P (1967) Field experience in estimating population growth. *Demography*, 4:228–243
- Little RJA (1988) Some statistical analysis issues at the World Fertility Survey. *The American Statistician* 42:31–36
- Madigan FC, Abernathy JR, Herrin A, Tan C (1976) Purposive concealment of death in household surveys in Misamis Oriental province. *Population Studies* 30:295–303
- Mahalanobis PC (1944) On large-scale sample surveys. *Philosophical Transactions of the Royal Society of London, Series B* 231:329–451
- Marks ES, Seltzer W, Krotki KJ (1974) *Population growth estimation: a handbook of vital statistics measurement*. The Population Council, New York
- Mayall J (1991) Non-intervention, self-determination and the “New World Order.” *International Affairs* 67:421–429
- McCormick JM, Mitchell N (1988) Is U.S. aid really linked to human rights in Latin America? *American Journal of Political Science* 32:231–239
- McDougal MS, Bebr G (1964) Human rights in the United Nations. *American Journal of International Law* 58:603–641
- McDougal MS, Lasswell HD, Chen L-C (1969) Human rights and world public order: a framework for policy-oriented inquiry. *American Journal of International Law* 63:237–269
- Mejia R (2006) Grim statistics. *Science* 301:288–291
- Metagora (2005) Pilot surveys on human rights issues (www.metagora.org/html/activities/act_hr.html accessed February 24, 2004)
- Mitchell NJ, McCormick JM (1988) Economic and political explanations of human rights violations. *World Politics: A Quarterly Journal of International Relations* 40:476–498
- Mitchell RE (1965a) Barriers to survey research in Asia and Latin America. *The American Behavioral Scientist* 9:6–12
- Mitchell RE (1965b) Survey materials collected in the developing countries: sampling, measurement, and interviewing obstacles to intra- and international comparisons. *International Social Science Journal* 17:665–685
- Murthy MN (1978) Use of sample surveys in national planning in developing countries. In Nambooditi NK (ed) *Survey sampling and measurement*. New York: Academic Press, pp 231–254
- Neier A (1989) Human rights in the Reagan era: acceptance in principle. *The Annals of the American Academy of Political and Social Science* 506:30–41
- Neuffer E (2001) *The key to my neighbor's house: seeking justice in Bosnia and Rwanda*. Picador, New York
- Neyman J (1934) On the two different aspects of the representation method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society* 97:558–606
- Nolde OF (1953) Human rights and the United Nations: appraisal and next steps. *Proceedings of the Academy of Political Science* 25:39–48
- Physicians for Human Rights (2001) *Endless brutality: war crimes in Chechnya*. Physicians for human rights, Boston
- Poe SC (1992) Human rights and economic aid allocation under Ronald Reagan and Jimmy Carter. *American Journal of Political Science* 36:147–167
- Poe SC, Meernick J (1995) US military aid in the 1980s: a global analysis. *Journal of Peace Research* 32:399–411

- Porpora DV (1990) *How holocausts happen: The United States in Central America*. Temple University Press, Philadelphia
- Regan PM (1995) U.S. economic aid and political repression: an empirical evaluation of U.S. foreign policy. *Political Research Quarterly* 48:613–628
- Reisman WM (1990) Sovereignty and human rights in contemporary international law. *American Journal of International Law* 84:866–876
- Roberts L, Lafta R, Garfield R, Khudhairi J, Burnham G (2004) Mortality before and after the 2003 invasion of Iraq: cluster sample survey (<http://image.thelancet.com/extras/04art10342web.pdf> accessed October 24, 2004)
- Robertson G (1999) *Crimes against humanity: the struggle for global justice*. The New Press, New York
- Sabagh G, Scott C (1967) A comparison of different survey techniques for obtaining vital data in a developing country. *Demography* 4:759–772
- Scheuren F (2002) Survey sampling technology in human rights research. *Amstat News* 306:77–78 (www.amstat.org/sections/ssoc/ accessed May 5, 2007)
- Scoble HM, Wiseberg LS (1974) Human rights and Amnesty International. *The Annals of the American Academy of Political and Social Science* 413:11–26
- Schwelb E (1970) The international protection of human rights: a survey of recent literature. *International Organization*, 24:74–92
- Sekar CC, Deming WE (1949) On a method of estimating birth and death rates and the extent of registration. *Journal of the American Statistical Association* 44:101–115
- Semb AJ (2000) The new practice of UN-authorized interventions: a slippery slope of forcible interference? *Journal of Peace Research* 37:469–488
- Shue H (1996) *Basic rights: subsistence, affluence, and U.S. foreign policy*, 2nd edn. Princeton University Press, Princeton
- Sikkink K (1998) Transnational politics, international relations theory, and human rights. *PS: Political Science and Politics* 31:516–523
- Silva R and Ball P (2006) The profile of human rights violations in Timor-Leste, 1974–1999: a report by the benetech human rights data analysis group to the commission on reception, Truth and Reconciliation of Timor-Leste. Benetech, Palo Alto
- Slater J, Nardin T (1986) Nonintervention and human rights. *The Journal of Politics* 48:86–96
- Sohn LB (1979) The improvement of the UN machinery on human rights. *International Studies Quarterly* 23:186–215
- Som RK (1997) Surveys on nutritional status of children in developing countries. *Proceedings of the Social Statistics Section of the American Statistical Association*, pp 170–174
- Spirer HF, Spierer L (1993) *Data analysis for monitoring human rights*. American Association for the Advancement of Science, Washington
- Stettinius Jr ER (1946) Human rights in the United Nations charter. *The Annals of the American Academy of Political and Social Science* 243:1–3
- Sudetic C (1998) *Blood and vengeance: one family's story of the war in Bosnia*. Penguin Books, New York
- Tezcan S, Omran AR (1981) Prevalence and reporting of induced abortion in Turkey; Two survey techniques. *Studies in Family Planning* 12:262–271
- Thakur R (1994) Human rights: Amnesty International and the United Nations. *Journal of Peace Research* 31:143–160
- United Nations (2005) *Household sample surveys in developing and transition countries*. United Nations, New York
- U.S. Commission on International Religious Freedom (USCIRF) (2005) *Report on asylum seekers in expedited removal*. USCIRF, Washington DC
- U.S. State Department Bureau of Democracy, Human Rights, and Labor, U.S. State Department Bureau of Intelligence and Research (2004) *Documenting atrocities in Darfur*. U.S. State Department, State Publication 11182, Washington
- Vaessen M (1978) World fertility survey: The current situation. *International Family Planning Perspectives and Digest* 4:73–74

- Verma V (1996) Comment on “Developing Samplers for Developing Countries.” *International Statistical Review* 64:156–162
- Verma V, Scott C, O’Muirheartaigh C (1980) Sample designs and sampling errors for the world fertility survey. *Journal of the Royal Statistical Society: Series A* 143:431–473
- Watkin K (2004) Controlling the use of force: a role for human rights norms in contemporary armed conflict. *American Journal of International Law* 98:1–34
- Weissbrodt D (1987) Humanitarian law in armed conflict: the role of international nongovernmental organizations. *Journal of Peace Research*, 24:297–306
- Werful F, Dunlop G (translator) (1934) *The 40 days of Musa Dagh*. Viking, New York
- Whiteman MM (1968) Mrs. Franklin D. Roosevelt and the Human Rights Commission. *American Journal of International Law* 62:918–921
- Yang D (1999) Convergence or divergence? Recent historical writings on the rape of Nanjing. *The American Historical Review* 104:842–865
- Yansaneh IS, Eltinge JL (2000) Design effect and cost issues for surveys in developing countries. Proceedings of the section on survey research methods of the American Statistical Association, pp 770–775

Chapter 2

The Statistics of Genocide

Mary W. Gray and Sharon Marek

Abstract In the aftermath of the Rwandan genocide, an attempt was made to use sampling to facilitate bringing alleged offenders to trial. This involvement led to the questioning of the role of statistics in the determination of whether killings can be characterized as genocide. Notwithstanding that the legal definition of genocide contains no reference to the magnitude of the offense, considerable attention has been focused on attempts at quantification. We review some prominent examples of alleged genocide and the techniques used to specify their magnitude.

“You’re taking a systematic sample,” remarked the trustee, distinguished looking in spite of his pink smock, the uniform of prisoners in Rwanda. This Harvard economics Ph.D. was the only one of the group surrounding us who grasped what we were doing in our rounds of the prisons of his country. Each prison governor was cordial, inviting us to see the improved conditions in their domain, pointing out that their charges were actually getting fat on the generous rations (albeit because they were getting little exercise, contrasted with the daily treks of 25 or more miles experienced by many previously in order to get water, food, and supplies, or to tend their fields). However, when we asked to see the prison records, in particular their lists of prisoners, we were met with puzzlement. Eventually, boxes of slips of paper were produced from chaotic cupboards or registry books were located.

What indeed was the purpose of that endeavor? At the time, 80,000 or more prisoners were in facilities designed for far fewer. International funding sources were becoming reluctant to build more prisons, having already added several sites. Each month as many as 10,000 new prisoners were entering the system while, with very few exceptions, death was the only way out. Two years after the 1994 massacre, the U.S. State Department’s Bureau of Democracy, Human Rights, and Labor was pressuring the government of Rwanda to bring prisoners to trial but was met with claims that the numbers involved were simply overwhelming. The Bureau’s creative solution was to begin by selecting a random sample of prisoners and bringing to trial those whose records would accommodate the prospect of fair trials under international law standards.

The concept was a good one and was originally signed on by the Rwandan government. The data collection proceeded, based on the concept of a sample stratified by region, as conditions across the country varied as the rescue forces

proceeded from east to west. One might expect, for example, better supporting documentation that might be used at trial as well as a different mix of the sorts of people arrested as the situation became less chaotic in the later stages. Although the records were ultimately produced, the information was generally very sparse; almost never were there names of arresting officers or the time and place of the alleged crime. Sometimes there was only the first name of a prisoner, other times only the family name—not, perhaps, that it mattered as we were told that many prisoners had given false names. Without more detail, it was clear that in the vast majority of cases the inadequate documentation and the passage of time would make it difficult to conduct a trial that would meet international standards of fairness.¹ However, what was present in nearly every record was the nature of the crime: “genocide.”

2.1 What is Genocide?

But was it genocide? The world, recognizing that it had done nothing to prevent the massive number of killings and generally accepting the collective guilt of its lack of response, named it so (Des Forges 1999). The horrendous nature of the killings—traces of which were still evident when we conducted the survey—and the huge magnitude of the tragedy are not in doubt. However, governments and human rights and humanitarian sources have estimated from 500,000 to 1,000,000 deaths, the variation giving a clear indication that it is unlikely that anything like an accurate figure will ever be known.

Article 2 of the Convention on the Prevention and Punishment of the Crime of Genocide states:

In the present Convention, genocide means any of the following acts committed with intent to destroy, in whole or in part, a national, ethnical, racial, or religious group, as such:

- Killing members of the group;
- Causing serious bodily or mental harm to members of the group;
- Deliberately inflicting on the group conditions of life calculated to bring about its physical destruction in whole or in part;
- Imposing measures intended to prevent births within the group;
- Forcibly transferring children of the group to another group.²

The usual discussion of genocide focuses on the issue of intent, not on the number of victims (Porter 1982). If there is no intent to destroy in whole or in part, widespread or systematic acts of persecution may still constitute a “crime

¹ Ultimately, the Rwandan government did not choose to work with the sample cases but rather proceeded, very slowly, to prosecute selected cases, chosen largely on the basis of the prominence of the potential defendant. From 1996 to 2005, 2,500 cases came before the courts. Beginning in March 2005, 56,000 are being tried in 118 Gacaca, semi-traditional community-based courts.

² The United Nations approved the Convention on December 9, 1948.



Fig. 2.1 Map showing genocide sites in Rwanda
Source: Yale University Genocide Studies Program.

against humanity.”³ There is no specific international convention dealing with crimes against humanity, but the International Criminal Tribunal for the Former Yugoslavia and the International Criminal Tribunal for Rwanda include this category in their statutes. Although definitions of crimes against humanity vary somewhat, what they have in common is the following:

1. They refer to specific acts of violence against persons whether these acts are committed in time of war or time of peace.
2. These acts must be the product of persecution against an identifiable group of persons irrespective of the make-up of that group or the purpose of the persecution (Bassiouni 1999).

We have thought of the massacre as being directed at the Tutsi minority population of Rwanda, the group traditionally favored by the Belgians in the colonial era to the detriment of the Hutus.⁴ This favoritism was perceived as continuing in the years

³ The term first appeared in the preamble of the 1907 Hague Convention, which codified the customary law of armed conflict.

⁴ In 1926 the Belgian colonial government introduced ethnic identity cards; anyone holding ten or more head of cattle was deemed a Tutsi (Orth 2004).

of independence, leading to a generally better socio-economic status for the Tutsis and an accumulation of grievances by the Hutus.⁵ Commentators differ on what motivated the mass slaughters, both on the part of the instigators and the thousands of “ordinary” people who killed their neighbors, Tutsi and Hutu (Mironko 2004). Some anthropologists and other experts have blamed the ethnic tensions on an arbitrary classification by the colonists. These authorities assert that in fact there was little difference in the two groups and that intermarriage tended to obscure much of what there may have been. Be that as it may, there came to be a deep resentment of the apparently better-off Tutsis on the part of many not-so-well-off Hutus—and who can forget the radio broadcast urging the extermination of the “cockroaches,” i.e. the Tutsis, recently highlighted by the popular film *Hotel Rwanda*? A history of inter-communal violence in Rwanda and neighboring Burundi—including fighting following incursions from Uganda by a mainly Tutsi exile group—further exacerbated enmities.

That the accepted version might not be the whole story first became apparent to us when a lawyer representing one of the defendants before the Arusha Tribunal⁶ contacted us to ask about how to determine how many of those killed were Tutsis and how many were Hutus. His client asserted that he was innocent of any role in the planning and execution of the killings, but even if he was involved, the massacres were not genocide. In that case, the Arusha Tribunal, an international forum for the prosecution of the participants in the genocide, would have no jurisdiction. He did not challenge the generally accepted range of magnitude of the massacres, but rather asserted that not ethnic origin, but rather economic status, was at the heart of the killings. People just decided, he claimed, that the best way to get the property of more prosperous neighbors was to kill them. It happened that the victims were disproportionately, but not exclusively, Tutsi because members of that group were disproportionately prosperous.⁷ That is, people were killed because their neighbors coveted their property, not because of their ethnicity, so the massacre did not meet the definition of genocide in the Convention.⁸ Of course, were the defendant guilty of “ordinary” murder, he could be brought to justice under Rwandan law which, unlike the Arusha Tribunal, provides for the death penalty. Perhaps he felt that he could elude the jurisdiction of Rwandan courts.

No matter how unattractive the defendant, our legal training as well as our personal inclination tells us that everyone is entitled to a defense. An examination

⁵ The government at the time of the massacre and some years previously was Hutu-dominated. The president tended to favor his home region, the predominantly Hutu northwest of the country, to the disadvantage of Hutus and Tutsis in other areas (Verwimp 2004).

⁶ The International Criminal Tribunal for Rwanda was established in Arusha in neighboring Tanzania.

⁷ It is generally agreed that some Hutus were killed because they attempted to assist their Tutsi friends and neighbors or by mistake. However, that the primary criterion for selecting victims was economic rather than ethnic appears largely speculative.

⁸ Some analysts have characterized the killings as President Habyarimana’s attempt to empower the peasant population, primarily Hutu, at the expense of the city dwellers, primarily although not exclusively Tutsi.

of the pre- and post-population records, given that ethnic origin had been recorded in Rwanda, might be used to determine whether his theory could be supported. However, this was bound to be a difficult and expensive process⁹ and would probably exceed our capabilities in French and the local Rwanda language as well as the time we had.

2.2 Do Numbers Matter?

Whether this defense was ever developed, we are unable to determine. But the experience did lead us to ask how the numbers of alleged genocide victims—and in some cases their race, ethnicity, or another factor establishing that mass killings were genocide—can be established. The Convention says nothing about numbers; realistically, however, killings generally only reach public attention once they are too numerous to be easily concealed. Although the Convention would require that only one or more persons be killed, the International Criminal Court has held that the conduct in question

- must have taken place in the context of a manifest pattern of similar conduct directed against the group, or
- was conduct that could itself effect destruction of the group.

What constitutes a “manifest pattern”? How widespread must the conduct be to “effect destruction”? Do 1,000 have to be victims? 10,000? 1,000,000? If only a few are killed, does this constitute the crime of “attempted genocide,” also punishable under the Convention?¹⁰ Even if a threshold level of destruction were to be recognized as essential for the characterization of killings as genocide, how do we then determine the actual extent of the killings? The circumstances surrounding such events obviously can make accurate counting a formidable task (Kuper 1981).

For example, how was the six million figure of victims of the Holocaust determined? How accurate are the widely differing estimates of Armenians killed by Turks made by Armenians and those made by Turks (who generally do not accept that there was genocide)? More recently, was Saddam Hussein responsible for killing—because of their ethnicity—enough Kurds or Marsh Arabs to justify a charge of genocide? Even though the Convention concerns itself only with the

⁹ Especially in view of the fact that the estimates of the total number killed vary so widely and that there has been no post-genocide census in Rwanda.

¹⁰ Article 3 states:

The following acts shall be punishable:

- (a) Genocide;
- (b) Conspiracy to commit genocide;
- (c) Direct and public incitement to commit genocide;
- (d) Attempt to commit genocide;
- (e) Complicity in genocide.

identity of the group being singled out and the intent of those responsible, not the numbers, the issue of how many is certain to arise in any prosecution.

The attempt to deal with the mass of prisoners in Rwanda also raises another question. Who should be charged with genocide?¹¹ Is driving a truck containing those who were involved in the killings sufficient?¹² Can soldiers or members of non-governmental forces be charged or is “just following orders” an adequate defense?¹³ What about those who claimed they killed only under coercion? How far the culpability for the Holocaust reaches remains unresolved although clearly not everyone who could be considered complicit has been charged in it or in more recent cases. Of course, many have been punished in some way without being formally charged, e.g., the “de-Nazification” process following World War II or more recently the “de-Bathification” process which has been sporadically carried out in Iraq. Or, as in the case of Rwanda, what are the numbers that a justice system can handle? How far down the chain of command can perpetrators be prosecuted or conversely, how far must one be removed from the actual genocidal acts to escape culpability, an issue now arising with respect to Saddam Hussein’s treatment of the Kurds?

As noted, one of the enduring controversies surrounding the killings in Rwanda is what motivated so many to participate in the killings. The political hierarchy may have been primarily concerned with consolidating Hutu power or maintaining a favored position for one region of the country or for an agricultural economy, but others may have been motivated by the promise or expectation of economic gain, the settling of old scores, or simply by coercion on the part of the Hutu Interahamwe local militia. Similarly, mixed motives could cloud the issue of whether other mass killings constitute genocide. And, of course, the demographic composition of the collection of victims can also be a complicating factor.

2.3 Historical Background

Genocide has a long and grim history (Durant 1950, 1957). Some disappearances of ancient civilizations may have been a result of genocide, although often combined with other factors such as poor management of the environment in the face of adverse conditions or “collateral damage” such as the inadvertent introduction of alien diseases into an indigenous population (Whitmore 1992). The term genocide is now being applied to other horrific events, e.g., the Inquisition, where thousands were executed because of their beliefs, or the Crusades, where thousands of Muslims, primarily civilians, were killed. Other massive exterminations of, e.g.,

¹¹ Article 4 states, Persons committing genocide or any of the other acts enumerated in article III shall be punished, whether they are constitutionally responsible rulers, public officials or private individuals.

¹² Under US criminal law, drivers of getaway vehicles waiting outside while their co-conspirators kill people inside are themselves guilty.

¹³ As a result of abuses in Iraq, US military law now makes clear that soldiers are obligated to refuse to follow an unlawful order.

indigenous populations in such countries as the United States (Stannard 1992), Mexico, or Australia may not meet the provisions of the Genocide Convention, but the difference is at best what would distinguish murder from criminally negligent homicide. But the magnitudes of these tragic events have been subject to widely varying estimates in part because the sizes of the original populations are subject to dispute.¹⁴

Other instances where estimates of the number of persons killed in genocidal incidents vary greatly include the massacres of Jews in Morocco in the eighth and eleventh centuries (Roumani 1976) and in Germany in the fourteenth century. In Europe the French slaughtered Huguenots and the Greeks killed Muslims. In their sweep the Mongols massacred many ethnic groups and in their empire the Ottomans killed Albanians, Armenians, Assyrians, Bulgars, Greeks, and Kurds, who in turn killed Christians, although whether all of these events could be characterized as genocides is not clear (Castellan 1992). The Iranians killed Baha'i in the nineteenth century, the Japanese killed Christians in the seventeenth century as did the Vietnamese in the nineteenth century. In the twelfth and thirteenth centuries, Hindus were massacred by Muslims, and no doubt many other instances could be cited.

Various incidents of massacres by colonial powers in Africa or between indigenous groups there both before and after the colonial period also most likely meet the definition of genocide, although even rough estimates of the magnitudes of the killings are difficult to obtain. The Germans in Southwest Africa (Wassink 2004) and the King of the Belgians in his Central African fiefdom can be added to the British in East Africa as perpetrators while regimes in Uganda, the Democratic Republic of the Congo, and Ethiopia have very likely engaged in genocidal conduct. Some see recent events in Zimbabwe as an eerie echo of the Khmer regime's assault on city dwellers. In Southeast Asia, the scale of the Indonesian atrocities in East Timor neared the 20% level experienced in Cambodia, although the base population was only 650,000.¹⁵

2.4 Calculations of Casualties

In his comprehensive series of studies of democide,¹⁶ Rudolph Rummel (1992, 1997) utilizes meta-analytic techniques, choosing usually to report from other authors a range of death estimates from low to high and then to put forward the figures he

¹⁴ For example, estimates of the Native North American population in 1500 range from about 1.5 million to as high as 12 million (Ubelaker 1992). That there were less than 250,000 Native Americans in the US in 1900 is fairly well agreed. Estimates of the aborigine population of Australia at the end of the eighteenth century are around 500,000, reduced to 50,000 by the end of the next century. The later figure is well-established, but the former is subject to dispute (Broome 1982).

¹⁵ The Zimbabwe and Cambodian regimes appear to have targeted urban dwellers and, especially in the case of Cambodia, the better educated populace.

¹⁶ A term invented by Rudolph Rummel. Under the definition of democide, Rummel includes genocide and other mass killings by states or other entities having control of territory, but not civilian and military war dead.

deems most credible.¹⁷ In particular, in most of his considerations he discounts outliers. However, this raises the question of whether one person's outlier is another's conservative estimate. Generally, as we would expect, the victim populations and/or those sympathetic to them present a high figure and the perpetrators one much lower, or the perpetrators do not admit the existence of a crime against humanity or even mass deaths if not genocide.¹⁸

But does the number of victims matter—particularly once some threshold is reached? Are the victims any less deserving of sympathy if their numbers are only in the hundreds rather than hundreds of thousands or the many millions of deaths attributed to Nazi Germany, the Soviet Union, Cambodia (Kiernan 2004), Japan, (James 1951), and China? Certainly not, but that does not mean that the numbers are unimportant. The media, while they often get the figures wrong, are obsessed with them. Often public interest is difficult to arouse unless an existing or impending tragedy is on an overwhelmingly large scale. Moreover, it is easy to conceal a few deaths, but more difficult to hide more extensive killing. Thus, it is worth looking at how figures are derived.

There are two basic methods one can employ: count the number of deaths or rely upon explaining a population deficit in terms of deaths (and other factors). The first requires better records of slaughter than are generally available,¹⁹ whereas the second relies upon being able to have reasonably accurate before and after population figures and a reliable estimate of expected population growth absent the genocide.²⁰ In general, demographic predictions are based on population growth in a period of time prior to the genocide in the population being studied, or subsequent rates in similar but unaffected populations.

¹⁷ Rummel asserts that the Nazis killed 5,291,000 Jews, 258,000 Roma, 10,547,000 Slavs, and 220,000 homosexuals (whose deaths would not qualify as genocide). Including hostages shot, reprisal murders, forced labor deaths, deaths from intentional famines, and political killings in addition to the genocidal murders, he arrives at a total of nearly 21 million victims of the Nazis.

¹⁸ The treatment by the Turkish government of the question of the deaths of Armenians demonstrates such denial, even though a predecessor regime was largely responsible (Dadrian 1986, Tashjian 1982). It should be noted that in the World War I period and the following few years, many Greeks, Nestorians, Kurds, and Christians were also killed by the Turks (Eagleton 1963).

¹⁹ One reason why the 6 million dead figure has come to be generally accepted is that the meticulousness of German records and their availability after the war have contributed to ease and rough agreement in making estimates of the number of victims. Nonetheless, the estimates of Jewish dead by the most reputable investigators vary by 40%.

²⁰ Such estimates are difficult. For example, in her recent study of the British treatment of the Kikuyu in Kenya in connection with the Mau Mau violence, Caroline Elkins estimates what the Kikuyu population would have been absent the mass murders and other harsh actions by the authorities. She relies on comparisons with the growth rate of nearby tribal groups. However, there is no comparison of the two groups in more tranquil times to authenticate this method. Elkins does present a convincing case for a number of deaths far exceeding the losses admitted by the British (Elkins 2005).

2.5 But Is It Genocide?

A complicating factor once the number of deaths is calculated as accurately as possible is to determine the source or motivation of the deaths. In many apparent genocidal situations, there are complex factors that may obscure the issue of whether there is genocidal intent, in particular in the fog of war (O'Balance 1973). For example, the main Armenian genocide occurred during a war that led to many civilian casualties as well as military losses (Sivard 1989). While of course soldiers often die because they are of a nationality, race, or ethnicity different from that of their opponents, their deaths are not considered genocide. The question of civilian casualties is more complex. Why were the atomic bombs dropped on Hiroshima and Nagasaki or why was Dresden firebombed? Was it to annihilate the enemy because of their nationality? Or rather to advance the cause of the victors in the name of a quicker end to the war and a limitation of eventual casualties? Histories are in the short run written by the victors, but revisionist histories may ultimately appear to present a different perspective.

The inadvertent introduction of disease has been mentioned as having a genocidal effect if not such an intent (Duplaix 1988). The imposition of occupations, blockades, or sanctions can also produce "collateral damage." However, such impositions differ from inadvertent actions because their intent is certainly somehow to damage the target population in order to further the perpetrators' goals.²¹

Currently, skeptics about the need for active United Nations or other external intervention question how many have died or been displaced in the Darfur crisis. There is also the question of whether the government is sufficiently involved and motivated by the intent requisite for the killings and displacement of people to constitute genocide. The media constantly seek a numerical basis for their reporting in Darfur. There are those who assert, in an echo of the claim of the Arusha defendant, that it is not their ethnicity that doomed the victims but rather that they had property which their neighbors coveted. It just happened, these commentators would aver, that those with property in the area where the Janjaweed operated were non-Arab villagers. The World Health Organization estimates that 70,000 died from disease and malnutrition in a six-month period in 2004; Amnesty International has estimated that 50,000 have been killed in the violence. However, the lack of cooperation from the Sudanese government has made compiling accurate figures very difficult. The UN High Commissioner for Refugees estimated that 200,000 sought refuge in Chad and 1.2 million were internally displaced, with 405 villages in Darfur having been completely destroyed from February 2003 through May 2005. The US government organized a project to conduct interviews of a systematic random sample of Sudanese refugees in Chad in part to determine the extent of the alleged crimes against humanity in Darfur and whether they could be characterized as genocide.

²¹ Whatever the original intent, the question of genocide may arise in the case of long-lasting severe deprivation.

In general the results of the interviews confirmed the participation in the attacks by the government of Sudan and their racial nature.²²

The massacre in Rwanda is not the only other case where the underlying criteria for choosing the victims are at issue (Saeedpour 1992). In the current discussion of the culpability of Saddam Hussein for the massacres of Kurds, three events are usually cited: the so-called “Barzani killings,” the 1988 gassing of the civilian population of Halabja, and the “Anfal,”²³ the massive killing and relocation of much of the Kurdish population of northern Iraq (Gendercide Watch 1988). While there might be general agreement that these constitute crimes against humanity, the question arises of whether the events, in particular the gassing, were “collateral damage” of the Iran–Iraq war.²⁴ Even if genocide is established, whether Saddam can be tied to the actions of those under his command may be problematic.

Leaving survivors leads to more accurate counting of the magnitude of mass killings. In 1983, Iraqi security troops rounded up men of the Barzani tribe, numbering 8,000 according to their families who were left behind, from resettlement camps near Erbil. All the men in this group, none of whom were known to be engaged in anti-government activities, were transported to southern Iraq and never heard of again. There were reports that some of the men were used as guinea pigs to test the effects of various chemical agents (van Bruinessen 1994). Just recently bodies of some of these men have been uncovered in southern Iraq by investigators including the Minister of Human Rights of the Kurdistan Regional Government.

This incident could be seen as a prelude to the “Anfal.” It might be maintained that the goal of the Anfal was merely to move the Kurdish population away from the Iranian border, not to eliminate the people. However, one written directive ordered that “all persons captured in those villages [in an area designated as ‘prohibited zones’] shall be detained and interrogated by the security service and those between the ages of 15 and 70 shall be executed after any useful information has been obtained from them” (Human Rights Watch 1995). An apocryphal story involves Ali Hassan al-Majid, now known as “Chemical Ali,” the Ba’ath party’s secretary general in charge of the “final solution” of what was perceived as “the Kurdish problem.” When confronted with the allegation that 182,000 Kurds had been killed, it is reported that he protested strongly that the deaths numbered 100,000 at most. In fact, the estimates of the number of victims of the Anfal are based on better evidence than many cases of alleged genocide. The women survivors, who generally were themselves deported to other areas of Iraqi Kurdistan, relate mass on-the-spot executions whose magnitude they can estimate with some accuracy, as they knew how many were in the targeted area. In other cases the entire populations of villages

²² Eighty-six percent of the interviewees reported joint Janjaweed–government attacks. A sizeable number of interviewees told of racial slurs on the part of the attackers (US Department of State 2005).

²³ Arabic for “The Spoils” from a Koranic reference to the treatment of unbelievers. The Anfal consisted of a series of killings, displacement, and destruction in 1987 and 1988 (al-Khalil 1989).

²⁴ Gulf War during 1980–1988. Many Kurds were thought to have supported Iranian forces against Saddam. Halabja, a town near the border with Iran, at one point before the gassing had been under joint Kurdish rebel–Iranian control (O’Leary 2002).

were slaughtered, but surviving records shown what the population figures had been. The 182,000 figure comes primarily from the fact that it is said that at least 1,200 villages were totally destroyed, perhaps as many as 2,000 including events outside the Anfal period itself. Reducing numbers derived from this information by estimates based on somewhat less reliable figures on refugees outside the borders and those in camps in other parts of Iraqi Kurdistan leads to the total estimate of fatalities.²⁵

The Khmer regime in Cambodia was responsible for deaths in the 1975–1979 period, estimated to total 1.7 million, over 20% of the country’s population. Religion was clearly the criterion for selection in the action against the many Buddhist monks who were slaughtered as well as for the attacks on Muslims and Christians, with ethnicity the reason for the killings of the Chinese, Vietnamese, and Thai minorities. However, a massive number of city dwellers, who tended to be better educated, were driven to the countryside where they were killed or died from starvation; 75% of the country’s teachers perished or fled abroad. There has been UN recognition that the actions of the Khmer Rouge constituted genocide, but as yet there is no legal accountability.

Reminiscent of the Nazi atrocities and the “Barzani killings” mentioned above, one of the most horrific events of the Bosnian genocide occurred in 1995 in Srebrenica, which had been designated as a safe haven under UN peacekeeping

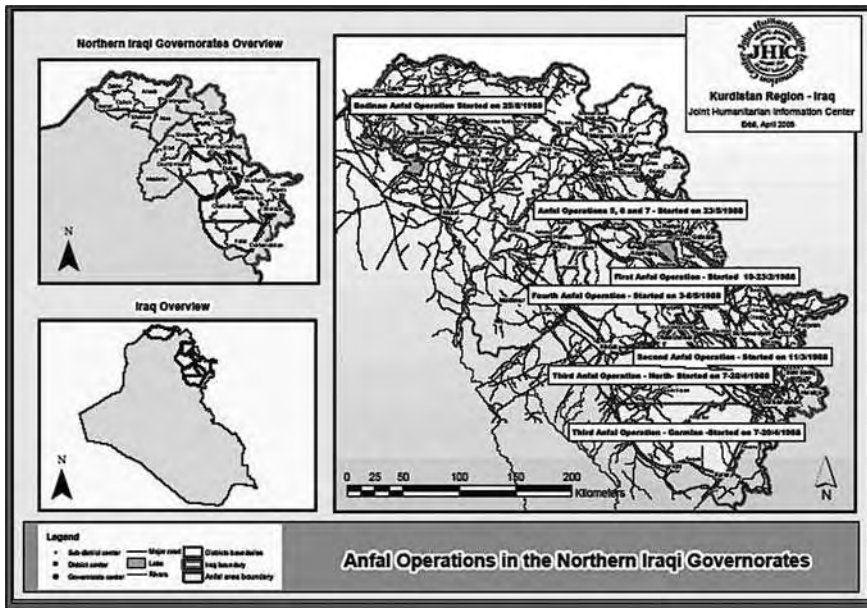


Fig. 2.2 Anfal operations in the northern Iraqi governorates
Source: Joint Humanitarian Information Center.

²⁵ During their post-Gulf War uprising, Kurdish forces were able to seize millions of documents from Iraqi archives documenting the extent and centralized direction of the Anfal.

supervision. Nearly 8,000 Muslim men and boys between the ages of 12 and 60 were systematically selected and slaughtered. The genocide of Bosnia did not end the violations in the Balkans. Atrocities were also committed a few years later in Kosovo. However, a UN-supervised court in Pristina ruled that in spite of a “systematic campaign . . . including murders, rapes, arsons and severe maltreatments” the actions of Serbian forces in Kosovo in 1998 and 1999 were not genocide as the purpose was “not the destruction of the Albanian ethnic group. . . but its forceful departure from Kosovo (BBC News 2001).”

2.6 Can Analysis Tell Us Why?

Many quantitative analyses of genocide also deal with the broader category of mass murders termed “democide.” Generally, there is no minimum number of such deaths required to include events in the data to be analyzed. Moreover, those types of death by governments include deaths through mistreatment of prisoners or forced laborers, killing of opponents of the government, famine, deportation, and non-combatants’ deaths during warfare (Gutman and Rieff 1999). Not surprisingly, the occurrences of these violations tend to be highly correlated with each other and with genocidal episodes as well. Factors that analysts consider include challenges to regime power; religious, racial, and ethnic diversity; ideological differences; minority stereotyping and discrimination; and the level of development in the country (Brogan 1989). Clearly, many of those are themselves related. As usual then, the question is how to model reality: regression (and if so, what kind), factor analysis, catastrophe theory (Rummel 1987)? Should the data be transformed, and if so, how? In his quantitative studies, Rummel (1997) has concluded that democide is basically incompatible with democracy, but others might disagree as to democracy’s effect. The claim that accountability will rein in the worst of government excesses may, however, be verifiable. Most analysts find that the extent of a regime’s power is a good indicator of the extent of any democide that might occur. Less unanimity surrounds the question of whether socio-economic factors are generally influential. However, there is agreement that they may in certain contexts.

2.7 Conclusions

The statistics of genocide are fuzzy at best and are likely to remain so. The use of GIS-equipped PDAs with aerial or satellite map imaging should increase the accuracy of baseline population figures.²⁶ Satellite imaging is also useful in identifying sites of the destruction of people or structures and should assist in estimates

²⁶ This technology was employed in the recent census in Timor Leste and is planned to be used in the upcoming census in Iraqi Kurdistan.

of the numbers of victims. More sophisticated techniques such as those employed in Kosovo (Ball and Asher 2002) may lead to more exact estimates in specific situations. Finally, better techniques for estimating population growth and decline, not addressed here, would contribute to our knowing with better precision the magnitude of past and, should we be unable to prevent them, future atrocities.

References

- al-Khalil S (1989) Republic of fear: the politics of modern Iraq. University of California Press, Berkeley
- Ball P, Asher J (2002) Statistics and Slobodan: using data analysis and statistics in the war crimes trial of former President Milošević. *Chance* 15:17–24
- Bassiouni MC (1999) Crimes against humanity. In: Gutman R and Rieff D (eds) *Crimes against humanity: what the public should know*. WW Norton & Company, New York, pp 107–108
- BBC News (2001) Kosovo assault “was not genocide” (<http://news.bbc.co.uk/1/hi/world/europe/1530781> June 22, 2005)
- Brogan P (1989) *World conflicts: why and where they are happening*. Bloomsbury, London
- Broome R (1982) *Aboriginal Australians: Black response to White dominance, 1788–1980*. George Allen & Unwin, Sydney
- Castellan G (1992) *History of the Balkans: from Mohammed the conqueror to Stalin*. Translated by Nicholas Bradley, Columbia University Press, New York
- Dadrian VN (1986) The Naim-Andonian documents on the World War I destruction of the Ottoman Armenians: the anatomy of genocide. *International Journal Middle East Studies* 18: 311–360
- Des Forges A (1999) *Leave none to tell the story: genocide in Rwanda*. Human Rights Watch, New York
- Duplaix N (1988) Fleas: The lethal leapers. *National Geographic* 173:672–694
- Durant W (1950) *The story of civilization: the age of faith*. Simon and Schuster, New York
- Durant W (1957) *The story of civilization: Part VI: The reformation*. Simon and Schuster, New York
- Eagleton Jr W (1963) *The Kurdish republic of 1946*. Oxford University Press, New York
- Elkins C (2005) *Imperial reckoning*. Henry Holt, New York
- Gendercide Watch (1988) Case study: The anfal campaign (Iraqi Kurdistan) (www.gendercide.org/case_anfal.html accessed June 22, 2005)
- Gutman R, Rieff D (eds) (1999) *Crimes of war: what the public should know*. WW Norton, New York
- Human Rights Watch (1995) *Iraq’s crime of genocide*. Human Rights Watch, New York
- James DH (1951) *The rise and fall of the Japanese empire*. George Allen & Unwin, London
- Kiernan B (2004) Recovering history and justice in Cambodia. *Comparative* 14:76–85
- Kuper L (1981) *Genocide: its political use in the twentieth century*. Yale University Press, New Haven
- Mironko CK (2004) Ibitero: means and motive in the Rwandan genocide. In: Cook SE (ed) *Genocide in Cambodia and Rwanda: new perspectives*. Yale Center for International and Area Studies Monograph Series No 1, New Haven, pp 173–199
- O’Balance E (1973) *The Kurdish revolt 1961–1970*. Archon Books, Hamden
- O’Leary C (2002) The Kurds of Iraq: recent history, future prospects. *Middle East Review of International Affairs Journal* 6:1–11
- Orth R (2004) Rwanda’s Hutu extremist insurgency: an eyewitness perspective. In: Cook SE (ed) *Genocide in Cambodia and Rwanda: New Perspectives*. Yale Center for International and Area Studies Monograph Series No 1, New Haven, pp 227–268

- Porter JN (1982) Introduction: what is genocide? Notes toward a definition. In: Porter JN (ed) *Genocide and human rights: a global anthology*. University Press of America, Washington, pp 2–32
- Roumani MM (1976) The case of the Jews from Arab countries: a neglected issue. In: Veenhoven WA (ed) *Case studies on human rights and fundamental freedoms: a world survey Vol 5*. Martinus Nijhoff Publishers, The Hague, pp 60–100
- Rummel RJ (1987) A catastrophe theory model of the conflict helix, with tests. *Behavioral Science* 32:241–266
- Rummel RJ (1992) *Democide: Nazi genocide and mass murder*. Transaction Publishers, New Brunswick
- Rummel RJ (1997) *Statistics of democide: genocide and mass murder since 1900*. Transaction Publishers, New Brunswick
- Saeedpour VB (1992) Establishing state motives for genocide: Iraq and the Kurds. In: Fein H (ed) *Genocide Watch*. Yale University Press, New Haven, pp 59–69
- Sivard RL (1989) *World military and social expenditures 1989*, 13th edn. Washington World Priorities, Washington
- Stannard DE (1992) Genocide in the Americas. *The Nation* 255:430–434
- Tashjian JH (1982) Genocide, the United Nations and the Armenians. In: Porter JN (ed) *Genocide and human rights: a global anthology*. University Press of America, Washington, pp 129–147
- Ubelaker DH (1992) North American Indian population size: changing perspectives. In: Verano JW and Ubelaker DH (eds) *Disease and demography in the Americas*. Smithsonian Institution Press, Washington, pp 169–176
- US Department of State (2005) *Documenting atrocities in Darfur*. Bureau of Democracy, Human Rights, and Labor and Bureau of Intelligence and Research, Washington
- van Bruinessen M (1994) Genocide in Kurdistan?: The suppression of the Dersim rebellion in Turkey (1937–38) and the chemical war against the Iraqi Kurds (1988). In: Andreopoulos GJ (ed) *Genocide: conceptual and historical dimensions*. University of Pennsylvania Press, Philadelphia, pp 156–157
- Verwimp P (2004) Peasant ideology and genocide in Rwanda under Habyarimana. In: Cook SE (ed) *Genocide in Cambodia and Rwanda: New perspectives*. Yale Center for International and Area Studies Monograph Series No 1, New Haven, pp 1–43
- Wassink J (2004) *Auf den Spuren des deutschen Völkermordes in Südwestafrika*. Martin Meidenbauer, München
- Whitmore TM (1992) *Disease and death in early colonial Mexico*. Westview Press, Boulder

Chapter 3

Why Estimate Direct and Indirect Casualties from War? The Rule of Proportionality and Casualty Estimates

Beth O. Daponte

Abstract This chapter considers the intersection of the Rule of Proportionality with demographic expertise. The concept of just war is reviewed and its application to prospectively choosing targets is examined. The author presents a framework for estimating excess deaths caused by war in a population, and discusses methodological and data issues surrounding such estimations. Work on the 1991 Persian Gulf War is presented to illustrate the application of the framework. Finally, the Rule of Proportionality is revisited, and a discussion of the application of demographic expertise and tools to the Rule of Proportionality are presented.

3.1 Introduction

With few exceptions, demographers have avoided applying their expertise in population dynamics and population estimation to making credible estimates of the mortality and morbidity impacts of conflict on populations. While there may be legitimate reluctance to venture into this area due to factors such as a lack of specific graduate-level courses offered that address estimating excess mortality from interventions and/or the lack of academic and professional freedom to make such estimates a professionally worthwhile endeavor, there are three reasons why demographers ought to involve themselves in this area.

First, in many cases, a lack of demographic involvement has left estimates of the impact of conflict to the realm of people with little expertise in measuring the impact of events on populations. They often provide the public with unsubstantiated “guesstimates” which are often made with incomplete information and through the prism of the value of life one places on the citizenry of each side of the conflict (which can lead to biases in the guesstimates). The guesstimates may be little more than back of the envelope calculations or “off the cuff” remarks on the possible magnitude of the impact of the conflict, but depending on the status of who has made the guesstimate, the figure may take on a life of its own and eventually be regarded as historical fact.

Second, and more importantly, the conduct of war is supposed to be dictated by the rule of war known as the Rule of Proportionality. The U.S. Air Force summarizes the rule, which originated with Clausewitz's work on war, in the following:

in conducting military operations, constant care must be taken to spare the civilian population, civilians, and civilian objects. With respect to attacks, the following precautions must be taken. Those who plan or decide upon an attack must: . . . take all feasible precautions in the choice of means and methods of attack with a view to avoiding, and in any event to minimizing, incidental loss of civilian life, injury to civilians, and damage to civilian objects; and refrain from deciding to launch any attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated (U.S. Department of the Air Force 1976).

In deciding on the legitimacy of potential targets, military strategists prospectively apply the Rule of Proportionality to various situations. In order to apply the rule, they must formulate an expectation of the number of noncombatant deaths that will result from attacking a target. That expectation is informed by knowledge of the impact on the population of past conflicts. Credible figures on the impact of past conflicts are a necessary element in applying the rule of proportionality.

Third, figures on the impact of conflict on populations can provide political scientists and theorists with information that will enable them to decide whether a war was or is being fought justly. Michael Walzer, in his book *Just and Unjust Wars*, writes of the morality of war. He writes, "war is always judged twice, first with reference to the reasons states have for fighting, secondly with reference to the means they adopt" (Walzer 1992, p. 21). Deciding whether a war is just can be distinguished from judging whether a war is fought justly. While political scientists, statesmen, and diplomats can debate the justness of the war, demographic expertise is needed to learn about the degree to which the war causes excess mortality amongst the population, whether subpopulations were uniquely targeted and, ultimately, the degree to which a war is or was fought justly.

A justly fought war abides by the Rule of Proportionality, incurring no more civilian deaths than necessary in gaining a concrete and direct military advantage from an attack. One can contrast this with Hitler's approach to war, where "The aim of war is not to reach definite lines but to annihilate the enemy physically" (Power 2003, p. 23). A war fought unjustly can be thought of as a free-for-all, whereas a war fought justly confines fighting to the military realm as much as possible. Just wars, according to Walzer, "are limited war; their conduct is governed by a set of rules designed to bar, so far as possible, the use of violence and coercion against noncombatant populations" (Walzer 1992, p. xvii). While one may question the morality of a given war, in theory an immoral war can be fought in a way that abides by the Rule of Proportionality.

Walzer asserts that retrospectively examining how a war was fought can be valuable in helping future wars to be fought justly. He writes, "... And even if the rules fail to shape the conduct of this war, they often succeed in shaping public judgments of conduct and so, perhaps, the training, commitment, and future conduct of soldiers" (Walzer 1992, p. xvii). That is, what demographers learn regarding the

mortality levels and patterns that resulted from a war can influence military policy at a later date.

In the remainder of this chapter, I present a framework for estimating excess deaths from war and its aftermath and discuss issues surrounding the data required for such estimates. Then, I summarize my work on the 1991 Persian Gulf War and revisit the Rule of Proportionality. I conclude with a discussion about how the demographic community can work together in bringing demographic expertise to this realm.

3.2 A Framework for Estimating War's Impact on Mortality

A demographic analysis of war and its aftermath is needed so that one can better understand the impact of interventions on the well-being of the population. In their examination of a conflict, demographers ought to be clear about what aspects of the conflict were examined and which were omitted from study, and what period was defined as "war." There can be differences of opinion on the start and end dates of a war—rarely do wars start and end with discrete time boundaries. The analyst needs to make clear which dates were used.

Fully examining the impact of a conflict means that the demographic analysis has examined the five possible ways that the war could have impacted the population. These five categories of excess death were first presented in Daponte (1993). Using the framework of the five categories of excess deaths, one can better understand and be explicit about what is known and remains unknown about the ultimate impact of a war on excess mortality in a population.

The framework considers only excess mortality from war, defined as mortality above and beyond what one would have expected in the population had demographic conditions and/or trends present just prior to the war or continued. The framework does not consider the total demographic impact of a war on a population, which would include, in addition to excess mortality, the fertility and migratory impacts of war.

In the short term, wars are known to decrease fertility and increase migration. While the impact on fertility and migration can change the trajectory of the population for years to come, the changes in fertility and migration usually result indirectly from conflict. Fertility is usually suppressed during the conflict and, often, baby booms follow conflicts. With respect to migration, while conflicts may produce refugee flows, often the majority of refugees return to their place of origin after fighting has ceased. While conflict usually impacts fertility and migration, the Rule of Proportionality does not speak of the fertility and migratory effects of war.

In the framework of excess mortality from conflict, the first category of excess deaths, military casualties from direct war effects, includes deaths to military personnel from the direct effects of the war. This category includes deaths to military personnel from shootings, bombings, missiles, and other direct forms of violence and assault.

The Rule of Proportionality does not discriminate between a large or small number of military casualties. Thus, the number of soldiers killed due to fighting is irrelevant to the rule's application since the military is seen as a legitimate target.¹ However, the number of military deaths on both sides of fighting is of tremendous interest to the military and general public both during and after the war.

The second category of excess death, civilian casualties from direct war effects, includes deaths of civilians from events directly associated with fighting. This category includes any incidents where civilians suffered violent deaths as a result of getting caught up in the war. Actions that result in civilian deaths include bombs and missiles missing their intended targets, misidentification of targets, civilians caught in cross-fire, civilians targeted in an effort to acquire territory, strafing of civilian populations, etc.

The third category of excess deaths, civilian casualties due to indirect war effects, includes deaths attributable to damage to the physical and public health infrastructure of a country and those due to refugee movements from the war. That is, deaths occur when the physical infrastructure (e.g., electrical power grids, water and sewage systems, fuel distribution systems and networks, and transportation infrastructure) is quickly taken from a population (Cuny 1983). This happens when the infrastructure is removed in the case of earthquakes and it happens when the infrastructure is removed in times of war. Civilian deaths due to postwar health effects can dwarf the number of civilian deaths due to war's direct effects since often a larger base of people are exposed to the former than the latter.

Further, compared with their mortality rates prior to fleeing, refugees often incur higher mortality rates due to a lack of access to physical and social infrastructure. In this framework, the number of excess deaths incurred among refugee groups are deaths attributable to war's indirect effects.

In the traditional interpretation of the Rule of Proportionality, only civilian casualties from direct war effects are balanced against the direct military advantage of an attack. However, I contend that civilian casualties from indirect war effects should also be included in this balancing equation, a point to which I return below.

The fourth and fifth categories of excess deaths, deaths due to the political chaos among (fourth) civilians and (fifth) military/government personnel after a war, should also be considered when tallying the war's cost. If a conquering nation leaves a war-torn area without a political and security infrastructure in place, and that instability provokes postwar chaos and violence (as was the case after the 1991 Persian Gulf War), then these deaths should also be considered when considering the extent to which the war caused excess casualties among the population. A war may leave a power vacuum, leading to political instability and postwar violence, resulting in excess deaths.

¹ Some (e.g., Walzer) have questioned the legitimacy of killing retreating, but not surrendering, soldiers.

3.3 Methodological Approach

3.3.1 *Civilian Casualties due to the Indirect Effects of Conflict*

Arriving at estimates of excess mortality in each of these categories presents substantial demographic challenges. Since estimating each of the categories of excess deaths may require different demographic approaches, each of the categories should be considered as a separate demographic question. The ultimate goal is to estimate a reasonable quantitative range of excess deaths. Optimally, one would want the age, sex, and ethnic distribution of each of the categories of excess mortality. Having such detail could help distinguish whether the excess mortality might be consistent with genocidal acts, as opposed to the incidental loss of civilian life.

Estimating the number of civilian casualties due to the indirect effects of war is essentially an exercise in running demographic projections multiple times using a variety of assumptions. That is, the counterfactual projection would start prior to the war and have built into it assumptions on the levels and patterns of fertility, mortality, and migration that would have prevailed had the war not occurred. Another set of projections would be run with the same fertility, mortality, and migration levels that were present prior to the war, then building into the projection the actual fertility, mortality, and migration levels that the population experienced during the war and the postwar period (time that it takes for the population to recover to its pre-war conditions). The difference between these sets of projections reflects the demographic indirect impact of the war.

There exist three challenges to creating such estimates:

1. *Isolating the impact of a war from other concurrent events.* It may not be possible to quantitatively disentangle the impact of a war from the impact of other punitive actions that may have concurrently occurred. For example, isolating the impact of the 1991 Persian Gulf War from the impact of UN sanctions on Iraq is impossible.
2. *The large uncertainty in such projections.* Point estimate projections, and even the traditional high, medium, and low projections, are inappropriate when comparing what the mortality experience of the population would have been had the war not occurred, and the projections of the population with war. Instead, a Bayesian Demographic approach (Daponte et al. 1997), which takes into account uncertainty in elements of the projection, should be taken for the estimation of expected and excess mortality.

In Bayesian Demographic projections, the base population and every component of the projection (fertility, mortality, and migration) are thought of as an uncertain quantity. The degree of uncertainty between the components of population change is allowed to differ. For example, rather than conceptualize a fertility rate as a number, the fertility rate is thought of as a range with a probability distribution function associated with it. One way a demographer may model the uncertainty in the fertility rate is, first, to review all of the data and other available information on fertility and then arrive at a mean of the total

fertility rate, and the 5% and 95% bounds on the total fertility rate. However, a normally distributed probability distribution function may not be appropriate, and the Bayesian Demographic approach allows one to model the rate using other types of probability distributions. By considering the uncertainty in a rate, one forms a credible interval of the component of the projection which makes the degree of uncertainty transparent to the consumers of the projection and other demographers.

In projecting of the population under the “with war” scenario, again, the base population and the components of the projection are thought of as uncertain quantities. The difference in the number of deaths between the two projection scenarios is one metric of the mortality impact of the war.

3. *Obtaining data necessary for the projections.* The projections require data on the population of the area under consideration before, during, and after the war. However, substantial technical and practical issues surface when collecting and analyzing such data. Optimally, what demographers would like is a situation where the age, sex, and race/ethnic distribution of an area’s population prior to the war is known, as are the demographic phenomena that prevailed. This generally requires that the demographer have available to him or her population data, usually from censuses and surveys.

There are problems with obtaining such data. First, even when data on the population prior to a war exist, such data may have moved into the realm of “sensitive” information and may no longer be available to demographers, other researchers, or the general public. Second, while aggregate data on the population prior to a war may be available from published reports, the level of detail needed may not be available. For example, one may be interested in particular regions of the country or in particular minority groups. There have been cases where such data has existed, but gaining access to the data in a form that would be useful after an area has erupted in violence was impossible. Such basic data becomes extremely “sensitive” to the country and to its foes during and after a war.

Third, obtaining estimates of mortality, migration, and fertility levels and patterns by demographic characteristics during and after the war is even more challenging. For these estimates to be available, there must be a party willing to sponsor the collection of such data. In a war-torn economy, donors may perceive that a country has more pressing needs than collecting data required for demographic estimation.

Fourth, even if the data are collected, the sampling properties used in the data collection may not be transparent, requiring that the demographer critically evaluate the usefulness and appropriateness of the data. Further, the data may only pertain to a particular point in time, and may not have external validity outside of the time frame or population sampled. This becomes relevant when situations are changing rapidly, as can be the case during wartime.

A lack of credible data can force a demographer who is interested in quantitatively estimating the impact of the war to rely on data of the quality that one would not usually use. This contributes to the discomfort that some demographers may have in producing estimates of excess mortality. The best that a

demographer may be able to produce are estimates derived from Bayesian population projections that have very wide credible intervals. Such estimates, though, can be informative when considering the Rule of Proportionality.

3.3.2 Military and Civilian Casualties from Direct War Effects

Considering military and civilian casualties from direct war effects, these estimates are usually the result of a tally of the number of people known to have died during various incidents. While theoretically these estimates can also result from sample surveys (where a sample of survivors are sampled, but also asked about people who may have died), what usually happens is that press reports or military reports of incidents of deaths are compiled to arrive at figures that represent a minimum number of deaths in these categories.

Estimating the direct effects of the war on the military is difficult and may not be a doable task. If there is a formal army or force, usually the force is aware of the number of deaths it has suffered. The issue becomes whether the government or force chooses to share such information with the public. In contrast, when there is an unorganized force involved in the conflict, the figures may be elusive. The force may be genuinely unaware of the totality of the losses it has suffered, or military or quasi-military deaths may be misclassified as civilian deaths.

One must use care when relying upon information on military and civilian casualties produced by one side of a conflict, since the information may contain biases. Each side of the conflict may overstate the damage it has inflicted on the enemy's forces and minimize the impact of the conflict on its own forces. It is also likely that the true losses suffered to the military may not surface until years after the conflict, when one might use figures based on census or administrative records to corroborate the estimates or claims made.

Considering civilians, the tally of deaths due to the direct effects of conflict can often be gleaned from compiling human rights and/or press reports. The risk in relying upon such reports is twofold: first, they may not encompass the entire universe of civilian deaths and, second, they may not represent an unduplicated accounting of deaths. Either error would bias the estimate, albeit in different directions and, of course, one cannot assume that these two types of errors balance. Similar to military deaths, each side may want to maximize or minimize the deaths that the enemy has inflicted, depending on whether the goal is to garner sympathy and/or show the cruelty of the enemy, or to maintain morale by demonstrating that the enemy has been unable to harm the population, and that the military has succeeded in protecting the homeland.

The degree of certainty needed about the number of civilian excess deaths due to direct war effects depends on the audience of the estimates. For example, the International Court of Justice would demand that the estimates of the civilian casualties from direct war effects include only deaths that can be thoroughly documented. This is in contrast to a more general audience, for whom a compilation of incidents from press reports might suffice.

3.3.3 *Military and Civilian Deaths due to Postwar Chaos*

To estimate the number of deaths due to postwar chaos one may tally the deaths, or be able to rely on a postwar, postchaos survey, depending on the magnitude of the chaos. Again, press reports, human rights reports, and government reports may exist on the situation. But in cases with suspected widespread deaths, sample surveys or comparisons of actual versus projected populations under a “no conflict” scenario years later may also be informative. The same issues discussed in the previous two sections arise—issues of uncertainty, issues of double-counting, issues of omission, etc. But theoretically, to estimate the complete impact of the war, one would need to take into account postwar chaos.

3.4 Applying the Framework to the 1991 Persian Gulf War

In estimating casualties from the 1991 Persian Gulf War (Daponte 1993), the above framework was used. Fortunately, there were a number of data sources on the Iraqi population prior to the war that were available that made it possible to demographically portray the country had the war not occurred. These data sources included

- the 1987 Iraqi census (which was available only in aggregate form and unavailable by ethnicity);
- the 1989 National Child Health Survey conducted by the Gulf Council of Health Ministers; and
- the 1990 Iraq Immunization, Maternal, and Childhood Mortality Survey conducted by the United Nations Children’s Fund (UNICEF).

Further, data on child mortality were collected shortly after the war, in August 1991, by the International Study Team. This ad hoc team consisted of approximately 50 volunteer researchers and had funding from UNICEF and the MacArthur foundation. Data on the individual level was made available for analysis, and the codebook for it was put on a napkin for me by one of the data collectors. The International Study Team did not make the data and a formal codebook for it available to a widespread audience, demonstrating the ad hoc nature of the organization of the data collection and dissemination. The data collected were vitally important, though, and without it one would not have been able to produce estimates of the indirect mortality impact of the war on the population in a timely manner.

Expected deaths could be estimated using the 1987 census and the 1989 and 1990 surveys. Deaths to the military during the war were estimated by William Arkin (1994), a military analyst, based on a reconstruction of bombings for which allied servicemen were given citations. Deaths to civilians during the war were gleaned from a database that I created from the books *Needless Deaths* (Middle East Watch 1991) and *War Crimes: A Report on United States War Crimes Against Iraq* (Clark 1992). Both books were compilations of witness reports of incidents that resulted in deaths from direct war effects. Further, the Iraqi government had released

statistics on civilian deaths due to the direct effects of war, by governorate. That is, deaths due to the direct effects of the war were estimated, essentially by tallying (see Daponte (1993) for details). Deaths to civilians due to the health effects of the war and other disruptions were estimated by creating model life tables using the information that was available on child mortality before, during, and after the war, and applying the various mortality levels to a projected population. Finally, civilian and military postwar deaths due to violence were based on US government sources and Middle East Watch (1992).

The findings breakdown as follows:

- military deaths due to direct war effects, 49,000–63,000;
- civilian deaths due to direct war effects, approximately 3,500;
- civilian deaths due to indirect war effects, approximately 111,000;
- military deaths due to postwar violence, approximately 5,000; and
- civilian deaths due to postwar violence, approximately 30,000.

In the 1991 Persian Gulf War, certainly, the largest category of death was that of civilian deaths due to indirect war effects. Many times more civilians died from indirect effects of the war than due to the war's direct effects.

The US military regarded the civilian infrastructure as a legitimate target. General Norman Schwarzkopf, commander of central command of the Allied Forces during the 1991 Gulf War, writes in his autobiography: "I presented [to Secretary of Defense Cheney] the detailed plans we'd worked up for air and sea strikes we could launch against Iraq. I ticked off lists of so-called high-value targets—military headquarters, power plants, factories—that we could quickly destroy" (Schwarzkopf and Petre 1992, p. 295).

Further, 3,500 civilians seemed to have perished due to the direct effects of the war. Whether this number violates the Rule of Proportionality is impossible to judge, because the direct military advantage of each of the attacks is not known. Ultimately, both the demographic information on excess mortality and the military information on the strategic advantage gained from an attack is needed to discern whether an attack was just. Demographers do not have any particular expertise that would allow them to apply the rule of proportionately. Their expertise is valuable only in calculating the impact of the conflict on the population, not in deciding the military advantage of an attack.

Was the Rule of Proportionality violated in the 1991 Persian Gulf War? Then and now, the Rule of Proportionality was and is not interpreted as including deaths due to war's indirect effects.

3.5 Revisiting the Rule of Proportionality

Given that conflict can affect a population for a period past the cessation of hostilities and that particular attacks are meant to avoid immediate deaths but inflict deaths at a later time, perhaps the Rule of Proportionality should be expanded in two important ways. First, the rule should include excess deaths due to postwar health

effects, preferably explicitly (as opposed to simply changing the interpretation of the rule). To illustrate why, consider the following scenario. Suppose a bomb hits a power plant and during the attack eight civilians die. The plant was destroyed, substantially weakening the enemy's strength. Such an incident may pass the traditional application of the Rule of Proportionality.²

However, suppose that the power generated by the plant was used to run the water purification system, upon which three million people relied for a stream of purified water. The attack of the plant had the ramification that the population no longer had access to potable water, which in turn increased disease rates and thus increased mortality rates, especially among children and the elderly. Suppose that 7,000 excess deaths occurred, and the lack of potable water was a contributing factor to those deaths. Then, under an expanded interpretation of the rule, the attack may have violated the Rule of Proportionality, depending on the concrete and direct military advantage gained.

Recent developments in weaponry (e.g., the development of "smart bombs") give the illusion of few casualties from bombing. While this is the case when considering civilian casualties from direct war effects, when one broadens the definition of casualties to include those who perish from the indirect effects of an attack, "smart bombs" may simply defer casualties from the proximate to the distal effects of an attack.

Days prior to the 2003 Iraq War, military personnel made statements implying that, when applying the Rule of Proportionality to the possible impending war in Iraq, they will not consider the long-term impact on civilians of destroying a "dual use" target (Lamuro 3/15/03). Rhetoric during the explicit fighting was that the civilian infrastructure was being spared. However, this rhetoric is at sharp contrast with the reality of Iraq's infrastructure during the most recent war.

When wars are fought so that civilian deaths from the direct effects are minimized but civilian deaths at a later date, due to infrastructure damage and other indirect causes, are not considered, then the international community needs to consider the implications. In particular, either the interpretation of the Rule of Proportionality or the rule itself should be updated so that the calculation is concrete and transparent, forcing the military advantage of an attack to be clearly balanced against both the direct and indirect civilian casualties.

Second, currently, the Rule of Proportionality does not apply to economic sanctions. Economic sanctions are a tool used to pressure a country into changing its conduct and are viewed as a less violent alternative to war, and often a substitute for war. Weighing sanctions against war, Elliott et al. wrote, "the key question is whether the price of sanctions would be higher than the economic and human costs of going to war soon" (1991, p. 259). Sanctions are a mechanism that a country or countries can use to harm another country without incurring any direct deaths, either to the imposing country(ies) or to the sanctioned country. But, as countries' economies have become more intertwined, the impact of economic sanctions can

² However, if the water purification system were targeted and destroyed, this would be in clear violation of the rules of war.

be devastating to a sanctioned population. As has been seen repeatedly (e.g., Cuba, South Africa) the direct military or even strategic advantage that the sanctions bring is unknown and, prospectively, uncertain.

What is the relationship of economic sanctions to the Rule of Proportionality? Even if the direct military advantage is unknown, demographic and epidemiological studies are needed to determine the impact of the sanctions on the population. To examine the impact on Iraqi children of economic sanctions imposed on Iraq prior to the 1991 Persian Gulf War, Daponte and Garfield used proportional hazards analysis, with the pre-war sanctions period indicated as time-varying covariate. The analysis showed that during the pre-war sanctions period, the risk of death to young children was substantially elevated (Daponte and Garfield 2000).

This raises two questions about the relationship of economic sanctions to the Rule of Proportionality:

1. Should economic sanctions fall under the purview of the Rule of Proportionality?
2. Do economic sanctions violate the Rule of Proportionality?

Based on the body of knowledge of the impact of economic sanctions on civilians and that sanctions are increasingly being used as a substitute for war, it seems that the Rule of Proportionality should be expanded to include this extension of the concept of war.

Because the rules of war have not kept pace with changes in the technology of war and economic warfare, which can minimize civilian deaths from direct war effects but yet have devastating indirect impacts on civilians, the rule needs to be rewritten so that it takes into account these changes in strategy.

3.6 Conclusions

Estimates of the excess mortality caused by the war should be part of the war's evaluation. In the United States, during and after each war, an evaluation of the war and the war's impact is done by the Department of Defense. The evaluations usually have minimal, if any, discussion of the impact of the war considered on the excess mortality of the population. A substantial part of the evaluations, though, are devoted to considering the equipment destroyed and damaged and casualties among US and strategic partners.

Any estimates of excess mortality will be suspect and controversial. Granted, some estimates are more thorough than others and rely on more realistic assumptions. Further, some estimates may have the aura of being "impartial," often depending on the credibility of their sponsorship. But many military officers and politicians would prefer that no estimates be made, making the argument of "Since we will never know with absolute certainty the number who have died, no estimates should be made." Estimating the human cost of war certainly reveals the short-term harm that occurs when diplomacy fails.

Refusing to produce estimates unless they concur with an unrealistic level of certainty is a way of hoping that if no estimate is produced, the question of excess

mortality becomes either irrelevant or the number of deaths becomes zero, at least in the public's mind.

With their expertise and tools at hand, demographers are able to make the most transparent and credible estimates of the impact that military ventures have on civilian populations. Because of the political nature of such estimates, any estimates must be transparent, perhaps beyond the transparency that many journals with space limitations allow. When making such estimates, demographers should make available the detailed technical notes on the estimations.

Excess deaths from war must be estimated, not solely to learn about past wars fought, but to take seriously the challenge that the Rule of Proportionality presents to demographers. A war that keeps civilians from harm's way has yet to be conducted. It is the moral obligation of demographers to apply their expertise to an issue that would allow policymakers to learn about the impact of strategies of past wars on populations.

The Rule of Proportionality can be thought of as the intersection of demography with the strategy and morality of war. The Rule of Proportionality demands demographic expertise to the same degree that democracy requires local population estimates.

Acknowledgment I thank participants in the IUSSP Seminar on the Demography of Conflict and Violence, November 2003, for their insightful comments on this work. I especially thank Helge Brunborg, William Seltzer, Ewa Tabeau, and Philip Verwimp for their comments. I also thank Joseph Kadane and Lara Wolfson. I am grateful to Herb Spirer and Arthur Spitzer's contributions to this work.

References

- Arkin W (1994) "Collateral damage" and Iraqi civilian deaths during Operation Desert Storm. Unpublished manuscript
- Clark K (1992) *War Crimes: A Report on U.S. War Crimes against Iraq*, Maisonneuve Press, Washington, DC.
- Cuny F (1983) *Disasters and development*. Oxford University Press, New York
- Daponte BO (1993) A case study in estimating casualties from war and its aftermath: The 1991 Persian Gulf War. *Physicians for Social Responsibility Quarterly* 3:57–66
- Daponte BO, Garfield R (2000) The effect of economic sanctions on the mortality of Iraqi children prior to the 1991 Persian Gulf War. *American Journal of Public Health* 90: 546–552
- Daponte BO, Kadane JB, Wolfson L (1997) Bayesian Demography: Projecting the Iraqi Kurdish population, 1977–1990. *Journal of the American Statistical Association* 92: 1256–1267
- Elliott K, Hufbauer G, Schott J (1991) Sanctions work: the historical record. In: Sifry M, Cerf C (eds) *The Gulf War readers: history, documents, opinions*. Times Books, New York
- Lamuro Col R (2003) National Public Radio, Weekend Edition. Interview (March 15, 2003)
- Middle East Watch (1991) *Needless deaths in the Gulf War: civilian casualties during the air campaign and violations of the laws of war*. Human Rights Watch, New York
- Middle East Watch (1992) *Endless torment: The 1991 uprising in Iraq and its aftermath*. Human Rights Watch, New York
- Power S (2003) *A problem from hell: America and the age of genocide*. Basic Books, New York
- Schwarzkopf Gen HN, Petre P (1992) *It doesn't take a hero: the autobiography*. Bantam Books, New York

- UNICEF, Regional Office for the Middle East & North Africa (1990) Iraq immunization, diarrhoeal disease, maternal, and childhood mortality survey, 1990, Evaluation Series No. 9. UNICEF, Cairo
- U.S. Department of the Air Force (1976) Air Force Pamphlet 110-31, International Law—The Conduct of Armed Conflict and Air Operations. U.S. Department of the Air Force, Washington, pp 5-9
- Walzer M (1992) Just and unjust wars: a moral argument with historical illustrations, 2nd edn. Basic Books, New York
- World Health Organization and UNICEF (1991) WHO/UNICEF special mission to Iraq, February 1991. Unpublished reports

Chapter 4

Statistical Thinking and Data Analysis: Enhancing Human Rights Work

Jorge L. Romeu

Abstract The role of statistical thinking and data analysis in human rights (HR) work goes beyond detection and standardization of HR violations and the exposure of violators. It includes the development of conceptual models to better understand such violations, as well as to obtain HR analyses of greater validity. This chapter overviews several statistical methodologies that have been successfully applied by the author in previous HR case studies. The author first discusses the data and their characteristics, including measurement scale levels, data collection, information sources, data quality, and their corresponding statistical problems. The author then discusses methods for comparing a case vs several controls, as well as for longitudinal analyses, using concomitant variables and information. Finally, he defines the concept of “differential increments” (the difference between HR violators’ and control countries’ performance measures), and applies it to comparing the multidimensional aspects of HR problems. Such a differential is used to assess the validity of some HR violators’ claims that HR violations are an unavoidable collateral *damage to social advancement*.

4.1 Problem Statement

The detection and assessment of human rights (HR) violations is not a trivial endeavor. On the contrary, it can be considered as an “open-ended” problem because the state of a country’s HR situation (as well as its socioeconomic development) is multidimensional. In addition, the problem components, as well as their relative importance (weights), are defined differently by different analysts. Let us explain that concept in detail.

Usually, there is more than one HR variable or factor in a problem (see the variables defined by Humana (1992)). Those variables can be grouped into three categories: political (e.g., the right to conduct free elections), economic (e.g., right to work and to receive a fair salary), and social (e.g., the right to an education and health care). Not all analysts agree that all the above are “human rights” and, thus, some are defined as individual rights and others as social rights. However, all categories should be taken into consideration if we want to build a consistent and widely

acceptable HR case, i.e. one that a majority will be willing to support and defend because it represents all the aspects (variables) of HR problems.

On the other hand, not all HR categories are simultaneously violated. Often, some HR categories are violated, while others are enhanced. For example, in Mao's China, in Stalin's USSR, and in Castro's Cuba, political parties were suppressed by a dictatorship that improved public education and health care for many of its citizens. Such situations may lead different analysts to different conclusions. For example, some analysts may consider only (or mainly) political variables, while others may consider only or mainly socioeconomic ones, and different analysts may weight the same variables differently. Hence, establishing and assessing adequate HR variables ends up being the first, and often the most difficult, part of a HR study.

In addition, seldom does the HR situation of one country (say X) completely dominate that of another country (say Y), for all p variables under study. In mathematical parlance, this is stated as $X_i \geq Y_i, i = 1 \dots p$. Hence, HR variable comparisons are performed through a (conscious or unconscious) process of vector dimension reduction, i.e., through the evaluation of a multivariate function " f " defined as

$$f(X) = \sum_{i=1}^p \alpha_i g_i(X_i) + \alpha_0 g_0(X_1, \dots, X_p) + \varepsilon,$$

In non-mathematical terms, such a function " f " weights each HR and socioeconomic variable with a single value $f(X)$. The α_i 's are the factor weights selected, the g_i 's are functions of the different HR variables included in the model, g_0 is a function that collects all possible variable interactions, and ε is the random error of the statistical model.

Analysts (mathematically or mentally) build such a model and assign weights to each variable to represent their relative importance in the HR and socioeconomic problem. For example, Stalinists believed that making education widely available to all was much more important (i.e., $\alpha_i = 10\alpha_j$) than holding free and open elections. The resulting mathematical expression $f(X)$ may not even be conscious for most analysts, but it is at the core of the discrepancies in HR data modeling and analysis.

As a result of all of the above, many HR analyses (e.g., analyses of China, Guatemala, Cuba) reflect the corresponding analysts' conscious or unconscious philosophies. This situation may help explain why analysts who have conflicting political affiliations often arrive at conflicting HR conclusions that are based on the same HR data.

Examples of such controversial HR assessments abound in the HR literature. Two recent such cases are those of the Cuban revolution of 1959, and the 30-year (1960–1990) civil war in Guatemala. In both cases, obvious HR violations were committed. However, very different results were obtained from the available data regarding HR violations, because of the different positions taken by HR analysts.

In the Cuban case, e.g., hundreds of university students and faculty have been expelled from or prevented from enrolling in the university. Thousands of others have been jailed, or have been fired from their jobs, for political reasons. And over a period of 48 years of one-party rule, one million citizens, 10 percent of the Cuban

population, have gone into exile. On the positive side, the Cuban government health and education policies have reduced illiteracy to practically zero, and have provided free and universal health care.

In Guatemala, during the long civil war, tens of thousands of civilians have died in obscure circumstances, or in overt massacres perpetrated by the army in their struggle against the guerrillas, and vice versa. On the other hand, after three decades of killings, the country has returned to a more or less pluralistic regime, where former guerrillas and military now jointly participate in public life instead of fighting an open war.

In both cases, supporters of these regimes justify as “unavoidable collateral damage” actions that opponents of the same regime qualify as HR violations. Both groups claim that “one cannot make an omelet without breaking some eggs.”

To complicate matters even more, HR statistics research stems from essentially two lines of work. One is provided by the highly qualified statistical experts that have become interested in human rights through readings, personal involvement, moral reasons, etc. This book includes examples of many of the most relevant exponents of that sector.

The second line of work stems from researchers such as this author, who have lived for years under regimes that systematically violate human rights. For example, this author was expelled from the University of Havana, then sent for over 2 years to the Cuban Unidades Militares para la Ayuda de Producción (UMAP) forced labor camps; he underwent political trials, was sentenced to jail, and has an extensive dossier with the political police, including detention, interrogation, and a suspended 12-year sentence, all over purely non-violent, political dissent activities, such as writing and publishing abroad a storybook about life in the UMAP camps. After arriving in exile, this author, like other HR workers, has used his statistics training and background to help assess HR situations in his country, as well as in others. The author’s HR case studies and work are referenced in the References list. This chapter is based on the methods implemented there and reflects the author’s past experiences.

Researchers from the above-mentioned second line of work may not always have as deep an understanding of the theory and methods of statistics as researchers from the first line. But they compensate for that with their first-hand experience in what we may call the HR violations body-of-knowledge: experience in HR violations that can never be obtained through readings, interviews, or visits to areas of interest to HR studies. And that experience gives those researchers a sixth sense about where to look, how to assess events, etc., that others lacking that experience may not have.

The author believes that this book on statistics and HR is greatly increased in practical value because it includes the work of researchers from both lines of work mentioned above. For both groups complement each other well and thus will provide unique contributions to the field of HR research. This book also demonstrates how good statistical thinking and modeling can create a more scientific context in which HR analyses can take place, where the inevitable trade-offs between different variable categories can be made under more equitable conditions, and where widely acceptable conclusions, leading to positive actions, can be obtained.

In the remainder of this chapter, the author develops three main analysis approaches. First, he discusses information sources, their measurement scale

levels, the characteristics of HR data, and the appropriateness of the statistical methodology employed in typical HR analysis. Then, the author discusses the concept of “differential increments” and, using paired comparisons (treatment vs control), he assesses HR violators via establishing a set of legitimate “control” countries. Finally, the author uses longitudinal studies, jointly with the graphical juxtaposition of qualitative concomitant variables, to build models that help explain the HR violations process, its origins, and some of its causes.

The author illustrates these methodologies and their use via several HR cases he has thoroughly studied previously (details of these Cuban case studies can be found in the original papers, referenced in the References list). However, the proposed analysis methodology can be extended or adapted to many other HR studies. The author illustrates that via his references to cases of HR violations in Guatemala during 30 years of civil war.

4.2 Examples of Data Problems in HR Studies

In addition to possible analyst bias, there are other non-trivial statistical problems that can affect HR data analyses. They have to do with the origin of the data, the data collectors, unit definitions, etc. All of them can weaken the resulting measurement scale level of the data and, hence, limit the statistical methodologies that can be appropriately used with them. Hence, just as we do with other statistics applications, we need to account for them carefully.

First, when examining HR data, we should be conscious of the type of political system we are assessing, the degree of control the violator country generating the data exerts, etc. In China under Chairman Mao, political and press control were very tight and few independent data were available. But this did not preclude using data from the UN and other international organizations about number of executions, political prisoners, etc. Also, anecdotal and TV coverage information on nationwide events such as the Tiananmen Square protests or the Cultural Revolution of the 1960s were available, however incomplete or otherwise faulty they might be. On the other hand, in a country where some free press, civil society, etc. still exists, the government is forced to allow greater degree of freedom and some independent data may be compiled, either directly or indirectly. This occurred under some Latin American dictatorships of the 1980s.

Also, we have the issue of the personnel that gather HR data. As stated above, HR sometimes becomes a football in partisan politics. Both sides of the argument (those interested in justifying, as well as those interested in exposing, HR violations) may exaggerate their counts or magnify the quality of actual counts by adding an “affective” dimension to the data. For example, data on the number of civilian casualties in a war may emphasize number of women and children killed, without including the context (e.g., that they may be used as human shields by one of the sides).

Another problem consists in establishing the threshold beyond which an internationally acceptable action, such as military interrogation of war prisoners, becomes

a HR violation (torture). That this is far from trivial is underlined by some recent interrogation procedures, which have been widely discussed, in conjunction with the legal definition of legitimate interrogation versus “torture” (see, e.g., Attorney General Alberto Gonzalez and the Abu Ghraib case, in *Newsweek International Special Investigation* www.msnbc.msn.com/id/4989481/ or the *Washington Monthly* www.washingtonmonthly.com/features/2004/0411.carter.html).

In addition, there are important data collection problems that arise when analyzing and comparing data from different countries. For example, when using UN yearbooks and third world countries’ censuses as sources, this researcher has found the following:

- different definitions for the same variables (does secondary education include normal or vocational schools?);
- different units (gross national product, given in domestic currencies);
- different time periods (results given per year versus per 5 years);
- overlapping periods (data collected from January to January vs from June to June);
- vanishing/appearing series (cost of living indices);
- changing bases (index numbers);
- changing definitions within a series (the value of the monetary unit in which, say, exports are reported, fluctuates from year to year); and
- biased, incomplete, or revised data.

Then, there are important effects caused by leaving certain concomitant variables out of the analysis of HR data. Some examples of variables excluded are

- the status of any pre-existing infrastructure (for it is not the same to increase literacy by 20 percent when the starting level was 10 percent than when it was 75 percent);
- consideration of the growth effects of the *S*-curve (for there is a steeper rate of growth in the middle of any process, when conditions have been established and needs have been discovered, than at its start or end);
- consideration of the saturation effect (for there is a limit to how many, say, miles of road should be constructed) and
- policy trade-offs to be made (20 kilometers more of roads versus a day care center) in the face of competition for limited resources.

All this will be further discussed in the section of this chapter on concomitant variables, where such information is used to show inconsistencies between a HR violator and a non-violator country.

Finally, a special situation occurs when comparing countries from different social and political regimes. In such comparisons, and to facilitate weight selection in the variable dimension reduction process, the use of “index numbers” has been suggested. However, the use of such indices to compare economic achievements of a HR violator with those of a non-violator, such as the Consumer Price Index (CPI), may prove somewhat controversial. Using the CPI would imply favoring the free-enterprise, consumer-oriented system over state-run economies. That would make difficult the fair comparison of nations with non-capitalistic systems, for example.

All the above problems contribute to weaken the measurement scale levels of the resulting HR variables, which may then be assumed to have an ordinal level. The author proposes that non-parametric methods be used when the above-described situations arise in HR data analysis. The author has had similar weak measurement scale experiences when analyzing hardware and software reliability data and has successfully used non-parametric statistical methods, with excellent results (Romeu et al. 2004; Romeu and Gloss-Soler 1983).

4.3 Characterization of “Differential Increments” in Country Comparisons

In hypothesis testing, we compare either with a standard (H_0 : the rate of death during the HR violations period is the same as an accepted overall rate) or with a peer (H_0 : rate of death during the HR violator’s control is the same as that of a non-HR violator, peer country). In HR work, it is often difficult to establish either of those two null hypotheses in a generally acceptable manner. Here, again, statistics can contribute to deal with the above-mentioned heterogeneity by using the proposed concept of “differential increments.”

A “differential increment” is the difference between the levels attained by a HR violator country, in some socioeconomic areas where it claims success, and similar socioeconomic achievements obtained by other comparable countries that are not HR violators. This approach seeks to show how similarly enabled countries can achieve similar success (be these the prevention of a dictatorship, or the rapid development of socio-economic levels) without having to resort to HR violations.

The author illustrates the derivation of such “differential increments” by comparing Cuban growth rates (used as the “case”), with those of three other “control” countries. Through this example he also shows how to implement that methodology for establishing “acceptable” control countries, a non-trivial but preliminary step in this approach.

The next section compares data from two adjacent, 40-year periods (before and after a claimed HR violation event), taken from Cuba (our case), and three other control countries, where such an “event” has not occurred. We first compare their socioeconomic data from the period of 1920–1960, to establish the four countries’ “similarity,” which will allow their use as controls. Then, we compare data from the period of 1960–2000, contrasting the growth rate of the case with that of the three controls.

4.4 An Example of the Case/Control Approach

If the HR problem were one of *testing a new drug*, our first concern as statisticians would be to find a *suitable* control. Here, the null (hypothesis) is that the treatment (HR violations) does not significantly increase the level of the response (better

socioeconomic parameters such as health care and education services) as compared with a non-HR-violator country.

However, we would not pair, say, a final stage, older cancer patient with a young, recently diagnosed one, to implement a clinical trial. We can also compare the case with itself used as its own control, by way of a *before and after* treatment approach. Here, the null (hypothesis) is that before the period of HR violations, the response in question (socioeconomic) was not significantly different than during the HR violations period.

However, finding a suitable control country is not easy. The author uses controls in two successive phases. First, he compares their indices to establish the similarity required to be a valid “control.” Then, he obtains the “differentials” in selected indices in order to assess whether the case (HR violator) has actually achieved higher than the (non-violator) control countries, as claimed.

To illustrate the author’s selection methodology, let us assume we want to use Guatemala as a control for the study of Cuba (the case). That would prove an inadequate choice. For even when size, population, and colonial history (both were Spanish colonies) are similar, they greatly differ in several important variables: ethnic composition (there are no Maya in Cuba, and few Africans in Guatemala), economic development, education levels, production systems, etc. (Cuba was much more advanced, economically and socially, than Guatemala.)

The Republic of Chile is a better match. Variables such as population size, socioeconomic, ethnic, religious, and educational indicators, and even its modern political history are more similar (including alternating periods of pluralism and dictatorship) to those of Cuba.

Three other potential controls are Costa Rica, Hungary, and Mexico. The first is one of the most stable and democratic nations in Latin America, but smaller and poorer than both Cuba and Chile. Hungary is an East European socialist country that recently evolved toward political/economic pluralism. It is similarly populated and has similar religious and socioeconomic features. But again, Hungary’s history and ethnic composition is different than Cuba’s and this may affect Hungary’s behavior and social philosophy. We therefore prefer Mexico to Hungary, ending with four Latin American countries that have a similar historic, economic, ethnic, and social background.

Chile is by far the closest match. Chile is larger in territory, but has a vast, thinly inhabited desert and frigid regions. Its population size is comparable to that of pre-/post-revolutionary Cuba, as are its racial composition, levels of education, and other socioeconomic variables.

Summarizing, Mexico is larger and Costa Rica is smaller than the case country in area and population. But both Mexico’s and Costa Rica’s political and economic systems have had strong similarities with that of Cuba, and the colonial histories of the three countries have many common points. For example, Spain trained many Mexican Viceroyas as Cuban governors. Even though Costa Rica is smaller than the case in size and population, it has also attained high socioeconomic standards. For many years Mexico, like Cuba, had a state-controlled economy and an hegemonic party political system. Costa Rica, in turn, has had a longstanding pluralistic one. Chile has experienced both pluralistic and authoritarian regimes in the recent past.

Cuba has had alternating periods of democratic governments and military regimes, followed by a long dictatorship with a state-controlled economy.

We have discussed the selection of these three control countries in the same spirit as we would discuss the selection of siblings to examine the effects of a treatment versus a control. Through the control countries, we can examine the null hypothesis that the case (HR violator) obtained similar gains as the controls in the responses selected. If we are unable to reject the null, the claim that HR violations are necessary collateral damage in an accelerated process of development is disproven.

In the same spirit that we spend time and effort in a clinical trial to validate the similarities between case and control, we need to spend (possibly much more) time and effort to validate the control countries selected. All details—about the countries selected as controls, about the reasons for specific socioeconomic indicators as variables, and about the case/control methodology discussed here—can be found in Romeu (1994,1996) and Romeu, Samuelson, and Spirer (1992).

4.4.1 An Example of Establishing the Validity of the Controls

In Table 4.1, we present data from The Statesman Yearbook (1929), which quantitatively support these four countries' strong similarities and illustrate our validation method. The selected variables provide a snapshot of their economic, political, and social conditions during the mid-1920s.

Variables selected include total population, population density, numbers of primary students, teachers, cattle, kilometers of paved roads, kilometers of railroad tracks, kilometers of telegraph wire, and post offices. These measurements are given either per capita, or per square kilometer.

Some reasons for selecting these specific variables include: (1) they reflect general education level, economic development, communications facilities, and (2) are available in the open literature. We do not expect all of them to be at par, in all countries. But their general socioeconomic picture, jointly with the historical-cultural-ethnic one, should point to such similarity. For example, we already showed

Table 4.1 Socioeconomic indicators in the mid-1920s

Variables	Cuba	Chile	Costa Rica	Mexico
Total population	3.57	3.75	0.45	14.9
Population density	31.05	4.9	20.5	19.7
Primary students	0.139	0.133	0.098	0.084
Teachers	0.00205	0.00288	0.00357	N/A
Cattle	1.337	0.511	0.919	0.375
Kilometers road	0.0234	0.0466	0.0056	N/A
Kilometers railroad Tracks	0.0684	0.0187	0.0179	0.0243
Kilometers telegraph lines	4.84	0.185	0.060	0.059
Post offices per capita	0.000221	0.000260	0.000718	0.000044

Source: The Statesman Yearbook (1929).

how Chile and Cuba have similar population sizes, coefficients for primary teachers and students, and numbers of post offices per capita.

A snapshot description such as that of Table 4.1 is insufficient to characterize a socioeconomic *process*. As we intend to compare these four countries not only at one instant but during a long period of time (over the twentieth century), we also need to assess the similarities between the *case* (Cuba) and the *controls* (Chile, Costa Rica, and Mexico) over the time series for these variables. We shall test the null (hypothesis) that the same *process* will continue past the threshold date of 1960, in all performance areas.

In Table 4.2, we show such data, taken at ten-year intervals: circa 1938, 1948, and 1958, respectively. The variables selected include population density (Dens), infant mortality (Mort), energy (Ener), primary students (Stud), and number of radio receivers (Rads) given either per capita or per thousand inhabitants. Those variables were selected because they reflect the levels of health, education, and nutrition in the population. Data are taken from the corresponding UN Yearbooks (1948 to 1993).

Notice, e.g., how the 1938 infant mortality rate for the case was comparable to that of 1958 for the three controls. This suggests that the case was historically more advanced in health care than most Latin American countries, including the three controls. In energy levels, the case started behind two of the controls, Mexico and Chile. But the case tripled its energy capacity during these 20 years, surpassing all three controls by the end of this period.

In primary education, Mexico made the most significant gains, while the others remained at constant levels and the case lagged behind. In number of radios (reflecting standard of living as well as level of public information) the case had the highest levels. Finally, illiteracy rates in 1950, which reflect basic education levels, were 20.6 percent (Costa Rica), 19.8 percent (Chill), 22.1 percent (Cuba), and 43.2 percent (Mexico). Only Mexico differed in this index, remaining way behind.

Table 4.2 Pre-1960 longitudinal comparison

Variable	Costa Rica			
	Chile	Rica	Cuba	Mexico
Dens (c. 1938)	6.41	11.29	38.1	9.51
Dens (c. 1948)	7.7	15.23	45.9	12.61
Dens (c. 1958)	9.84	21.1	56.5	16.43
Mort (c. 1938)	235.7	123.1	83.0	128.0
Mort (c. 1948)	160.4	93.3	N/A	99.7
Mort (c. 1958)	126.8	89.0	34.7	80.8
Ener (c. 1938)	0.67	0.17	0.34	0.44
Ener (c. 1948)	0.76	0.22	0.47	0.61
Ener (c. 1958)	0.80	0.26	0.93	0.75
Stud (c. 1948)	0.131	0.144	0.107	0.116
Stud (c. 1958)	0.138	0.156	0.108	0.147
Rads (c. 1938)	0.031	N/A	0.034	0.019
Rads (c. 1948)	0.096	0.029	0.109	0.030
Rads (c. 1958)	0.089	0.070	0.170	0.077

Source: United Nations Yearbooks (1948–1993).

In the first section of this chapter, we discussed how “learning curves” and pre-conditions should be taken into consideration (as concomitant variables) to assess the case/control relative growth rates and their socioeconomic gains. Our indicators were obtained from the initial phase, where case and controls are established as relatively homogeneous, as part of the process of preparing a valid case/control country comparison for the second phase (1960–2000).

4.4.2 Example of HR Violation Assessments via Pre- /Post-Test Comparisons

The initial step in our before/after comparison consists of selecting the data. In the present example, those come for the periods before and after 1960 (sources include Oficina Nacional de Estadísticas (1971, 1981); Oficina Nacional de los Censos Demográfico y Electoral de Cuba (1954–1955); Eberstadt (1986); Gordon (1983); Mesa Lago (1971, 1981, 1987, 2000). With those data, we assess the null hypothesis that improvement levels obtained by the case during the HR violations period are not significantly different than those for the controls, and therefore cannot be used to justify the HR violations as collateral damage associated with such socioeconomic improvements.

Results from Cuba’s 1953, 1971, and 1980 censuses yield that the percent of the workforce employed by the government was, respectively, 8.8 percent, 87 percent, and 93 percent. There were 740,000 private workers reported in 1962, but only 175,000 in 1978. During that period, however, the population grew from 6 to 8 million. Hence, these data suggest a shift in type of employment patterns, rather than actual employment growth.

In the realm of public education, the 1953 census reports an adult literacy rate of 73 percent. This value went up to about 96 percent after massive literacy campaigns in the early 1960s. There were only three public universities with 25,000 students and 2,500 professors in 1956. In 1986, there were over a dozen universities, with 256,000 third cycle students. This shows an overwhelming increase in the levels of education, for the country, during the violations period.

On the other hand, there were massive politically and religiously motivated faculty and student purges in the 1960s. And, until the mid-1980s, openly religious students were banned from registering for careers in medicine, economics, engineering, journalism, and others, in a country that, in 1956, was 80 percent self-avowed Catholics and 8 percent Protestants.

In the area of health care, Cuba went from a life expectancy of 64 years (1960) to 74.2 (1984); from an infant mortality rate of 34.7 per 1,000 (1959) to 10.2 (1992); from 0.93 physicians per 1,000 people (1959) to 4.33 (1992); from 0.74 nurses per 1,000 people (1959) to 6.83 (1992); from 4.22 hospital beds per 1,000 people (1959) to 6.1 (1992)—Alonso and Lago (1994). Those data show a significant increase in health care levels during the HR violations period.

Internal migration is constrained because the government controls both housing and food ration cards. The 1980 Census reports an average growth, among the 14

provincial capitals, of 17 percent (the country grew 25 percent), and Havana shows only 7.7 percent growth, half that of any other. Emigration is also severely controlled by the government; in spite of that, 10 percent of the population has left since 1960. Over 120,000 people left during the 1980 Mariel Boatlift and 35,000 more during the 1994 Raft Exodus. Exiles include tens of thousands of university professionals, administrators and technicians, forever lost to contributing toward the country's advance.

HR violations in Cuba are well documented and have occurred as a result of (i) peaceful political or religious activism, or for exercising the right to (ii) economic freedom or (iii) free emigration, when (i) and (ii) were not possible. The claim that such HR violations constitute an unavoidable "collateral damage" in the process of obtaining a larger "good" is often used as a defense by violator supporters. As discussed in our initial section, this is a classic multi-dimensional problem where there are significant gains in some areas, and significant losses in others. It is a matter of which ones the analyst chooses to report, how they are reported, and what kinds of trade-offs and valuations of the variables in question are performed.

4.4.3 An Example of the Use of "Differential Increments"

Another way of assessing the claims previously described consists in comparing socioeconomic statistics of the case with similar ones obtained from the three controls, for the same time periods, using the "differential increments" approach.

For example, in the previous section we showed how the case's pre-1960 economic indicators were often better than those of the three controls. It could be reasonably conjectured that such pattern would continue for the following period. It would thus be reasonable to apply a flat 4 percent yearly growth rate to the case, and compare such long-term forecasts to the actual values, for the 40-year period of 1960–2000. Any difference between both results could be attributed to the effects of the case's policies.

However, this approach is questionable, given that the world significantly changed in the second half of the twentieth century with respect to the first half. There were new technical, geopolitical, and economic factors that did not exist before. Any valid time series analysis is based on the stability of the underlying process, which may not exist here. Hence, we shall not pursue this approach.

Instead, we shall use the actual levels attained by the three above selected control countries, whose 1920–1960 growth rates were at par or below those of Cuba, for the same period. We will use them to compare the case of actual achievements with what it could presumably have obtained without the policies that led to the claimed HR violations. The comparison is valid, for both case and controls underwent, in the second period, the same changes and other factors. The difference between their achievements would then provide the *differential increments*.

We present a longitudinal study for the time period after 1960 in Table 4.3, and a snapshot of the socioeconomic conditions of these four similar countries, at the end of the 1960–2000 period, in Table 4.4

Table 4.3 Post-1960 longitudinal comparison

Variables	Chile	Costa Rica	Cuba	Mexico
Dens (1970)	13.2	34.1	73.3	24.9
Dens (1980)	14.9	44.0	86.3	36.5
Dens (1990)	17.7	58.7	92.6	43.7
Mort (1980)	47	30	22	58
Mort (1990)	18	16	15	41
FLif (1980)	70.6	73.1	74.8	68.4
FLif (1990)	75.1	77.7	77.0	72.1
Ener (1970)	86	67	74	66
Ener (1980)	135	145	150	155
Ener (1986)	170	193	200	221
StdN (1975)	0.224	0.183	0.192	0.190
StdN (1980)	0.197	0.155	0.148	0.204
StdN (1989)	0.151	0.141	0.083	0.168
Rads (1975)	164	77	194	111
Rads (1985)	330	246	326	189
Rads (1990)	340	259	343	242

Source: United Nations Yearbooks (1949–1993)

Table 4.3 includes population density (Dens), infant mortality (Mort), female life expectancy (FLif), energy consumption (Ener), primary students (StdN), and radio receivers (Rads), per capita or per thousands.

From Table 4.3, we see how Cuba and Chile have contained population growth, an indicator of social advance, as opposed to Mexico (who has only slowed it). Next, we see how Chile and Costa Rica have gone from a level of infant mortality much higher than Cuba to one comparable to it. Mexico also made advances, but still lags behind in this indicator. In life expectancy, however, all three control countries have attained much higher levels, and Costa Rica has matched those of the case. These health indicators signal how there have also been large health improvements brought by new technical and medical advances independently of political regimes throughout Latin America (and the world).

As the null hypothesis of no difference is not rejected, it is then possible to conjecture that the case's larger advances in health care would have been obtained anyway, especially when Cuba already had, since the 1950s, higher health standards and a longer health care tradition.

In energy consumption (an indicator of industrial development and general standards of living), Costa Rica maintained a growth level similar to that of Cuba, but Mexico had a larger growth rate and Chile a smaller growth rate than Cuba. Primary students have decreased as a percentage of the general population. But this has been a general phenomenon in developing countries. Higher and technical education level percentages continue to be larger in Cuba. But the three controls are rapidly closing the gap. Finally, the number of radio receivers is now similar in Cuba and Chile while Costa Rica and Mexico are lagging behind.

Table 4.3 allows us to compare, not only the *level* attained by each country, but also its *growth rate*. Mexico is still the one with the highest population growth and

lowest socioeconomic indices among these four countries. Notice how Chile, Costa Rica, and Cuba departed from different levels (Cuba is generally ahead). However, all four have practically closed the gap by the end of the twentieth century.

Table 4.4 shows some 1990s United Nations statistics. Notice the close agreement in indicators from Cuba, Chile, and Costa Rica, showing how significant socioeconomic advances have been achieved by all countries in the region during the last 40 years, independent of the type of socioeconomic system that prevails in each.

Table 4.4 also shows how Chile and Costa Rica reduced their illiteracy rate to single digits, and Mexico from 44 percent to 17 percent. With regard to intake of calories and proteins (nutrition indicators), Cuba is still about 10 percent above the three control countries in the former, and about the same in the latter. However, given the pre-1960 high health standards of the case, this is no surprise.

In energy consumption, the basis of industrialization, the case is positioned between controls Mexico and Chile. Costa Rica, a mainly agricultural and rural nation, lags behind. Finally, in standard of living indicators (phones, TV, radios), the case levels of pre-1960 years have been surpassed or attained by the control countries. Given its higher levels in the 1950s, it seems that a higher growth rate was achieved by the three controls with regard to standards of living. But we would still need to consider the saturation effect of the “S-curve.”

We now show how we estimate the *differential increments*, with respect to Costa Rica, for the variable *infant mortality*. The differential is 2.8 (per 1,000). With respect to literacy, the differential is 3.6 percent. Those two values are obtained by assuming that Cuba would have attained at least the illiteracy rate of the control Costa Rica (7.4 percent) using the pre-1960 model, instead of its own value of 3.8 percent.

This premise is based on the fact that, under the previous model, Cuba was always at par or above Costa Rica in educational indicators. Therefore, any

Table 4.4 Socioeconomic indicators in the 1990s

Variables	Chile	Costa Rica	Cuba	Mexico
Illiteracy (per 100 people)	8.9	7.4	3.8	17.0
Infant mortality (per 1,000 infants)	17.1	13.9	11.1	43.0
Expected male life span (years)	68.1	72.4	72.6	62.1
Expected female life span (years)	75.1	77.0	76.1	66.0
Calorie intake (daily average)	2,480	2,711	3,153	2,986
Protein intake (daily average)	69.6	64	71.6	81.5
Cement (000s Tons)	2,115	N/A	3,696	24,683
Energy (000s KW)	1,270	602	1,461	1,788
Phones (per 1,000 people)	8.3	14.9	5.8	11.8
TV sets (per 1,000 people)	201	136	203	127
Radios (per 1,000 people)	340	259	343	242
Newspapers (units)	47	6	15	216
Students-1 (000s)	1,991	422	885	14,508
Students-2 (000s)	742	123	1,073	6,704
Population (per square kilometer)	13.1	3.0	10.6	86.2

Source: United Nations Yearbooks 1993

difference between the two may be attributed to the effect of the current Cuban socioeconomic development model. This would constitute a legitimate estimate of the *differential* for the variable *eradication of illiteracy*. All other differentials can be obtained in a similar manner.

Another approach would consist in using non-parametric regression to estimate the *differentials*. For illustration, we show the percent indices of total production of electricity for 1970–1986 in Table 4.5, taken from Wilkie (1990). The year 1975 corresponds to 100 percent. The four-country average values (per year) are also given. In Table 4.5 we have also included the slope and Index of Fit (IoF) of the parametric linear regression, and the slope (NPSlp) and confidence interval (CI) for the non-parametric slope obtained from these data. Regressions were first obtained separately for each country, then for country averages, and finally for the combined three controls. Hence, the combined regression column corresponds to the ($3 \times 4 = 12$) control data points. We obtain the parametric and non-parametric slopes (8.02 and 8.5) from the combined regression, and use them to estimate the general growth for electrical power variable. We then compare them with the growth rates (slopes) for Cuba alone (7.7 and 7.5). Using the slope of the combined data, we obtain a higher index for the 1985 electricity production than the case actually achieved.

We also obtain a 90 percent non-parametric confidence interval for the slope of the combined data, and note the case slope is included within that confidence interval. We cannot reject the null hypothesis that the case growth in electricity production during the 1960–2000 period is similar to that of the combined three controls, none of which followed the case's development model (which involved violating HR).

This analysis supports the view that regional growth, in general, has been similar in countries with different socioeconomic systems, for all have caught up with the case's growth rates. Such results lead us to question the case's claim that to obtain higher achievements in health and education HR violations were unavoidable.

This argument is not unique to Cuba. The same has been used in Guatemala, with respect to the guerrilla insurrection. Costa Rica and Honduras, e.g., two regional and similarly small and developing countries, did not resort to Guatemala's levels of violence and were not overrun by guerrilla insurrection either.

Table 4.5 Analysis of total electricity production: 1975 = 100%

Countries	1970	1975	1980	1985	Slope	IoF	NPSlp	CI
Chile	86	100	135	161	5.2	97	5.1	4.9–6.1
Costa Rica	67	100	202	185	9.1	81	8.2	6.6–13.5
Cuba	74	100	150	185	7.7	98	7.5	7.0–8.5
Mexico	66	100	155	216	10.1	98	10.5	8.9–11.6
Average	73	100	164	187	8.04	97	8.1	5.5–8.7
Combined					8.02	86	8.5	6.9–11.0

Source: Wilkie (1990)

4.5 An Example of the Longitudinal Study Approach

This methodology consists of modeling, both graphically and analytically, the time series of HR violations within its context, in the manner physicians do with an electrocardiogram during a stress test, or in the manner the advance and retreat of Napoleon's forces were contrasted with number of casualties and dates, during his invasion of Russia in 1812, on a milestone statistical chart (see www.edwardtufte.com/tufte/posters). First, we look at the changes and anomalies in the time series pattern. Then, we associate those with the changes in their contextual conditions, and try to infer causes and effects. Let us illustrate this approach with an example.

Let the Cuban Gross Domestic Product (GDP) for 1980–2000 be modeled as the time series $\{X_t\}$ that depends on social domestic policies, Soviet and Western European subsidies and businesses, and the effect of the US economic embargo. In mathematical parlance we write,

$$X_t = f(g_1(W_t), g_2(Y), g_3(Z_t), g_4(W, Y, Z)) + E_t$$

where $E_t \sim F(1980 \leq t \leq 2000)$, and with the factors:

- i. W_t , the Case economic/social domestic policies,
- ii. Y_t , Soviet and Western European subsidies and businesses, and
- iii. Z_t , effect of the US economic embargo.

As usual, E_t is the statistical model error term, distributed F , and g_4 collects all possible interactions between the model variables.

This conceptual model helps understand and assess the previously mentioned claims, made by Cuba, about its HR violations and its causes, by relating them to the overall 1980–2000 socioeconomic period (Fig. 4.1), as well as to their historical context. Let us give the background of such context.

In 1980, 10,000 persons, in 48 hours, sought asylum in the Peruvian Embassy in Havana. Shortly after, 125,000 people sailed to the United States via the Mariel Boatlift. After these events, the Cuban government allowed some forms of private enterprise and free markets, up to then totally forbidden in Cuba. As a result, the economy grew considerably in the 1981–1985 period. In 1986, during the period of “Rectification of Errors,” the Cuban government again prohibited most forms of private enterprise, and the economy shrank again.

At the end of the 1980s, internal struggles in the communist party led to the trial and execution of General A. Ochoa and other military officers as well as to an increase in the internal dissidence movement. Then came, in 1989, the Glastnost and Perestroika movements, followed by the break-up of the Soviet Union, the end of the Cold War, and the loss of all Soviet subsidies: the Cuban economy collapsed.

The Cuban government then reversed its traditional economic policy. It established joint ventures with large European tourism companies, starting a new “state

capitalism” that saved the economic situation but exacerbated the internal dissident movement.

As a result of the large influx of foreign tourists, an explosion in male and female prostitution, drug use, etc. reappeared, all of which had disappeared from Cuba since

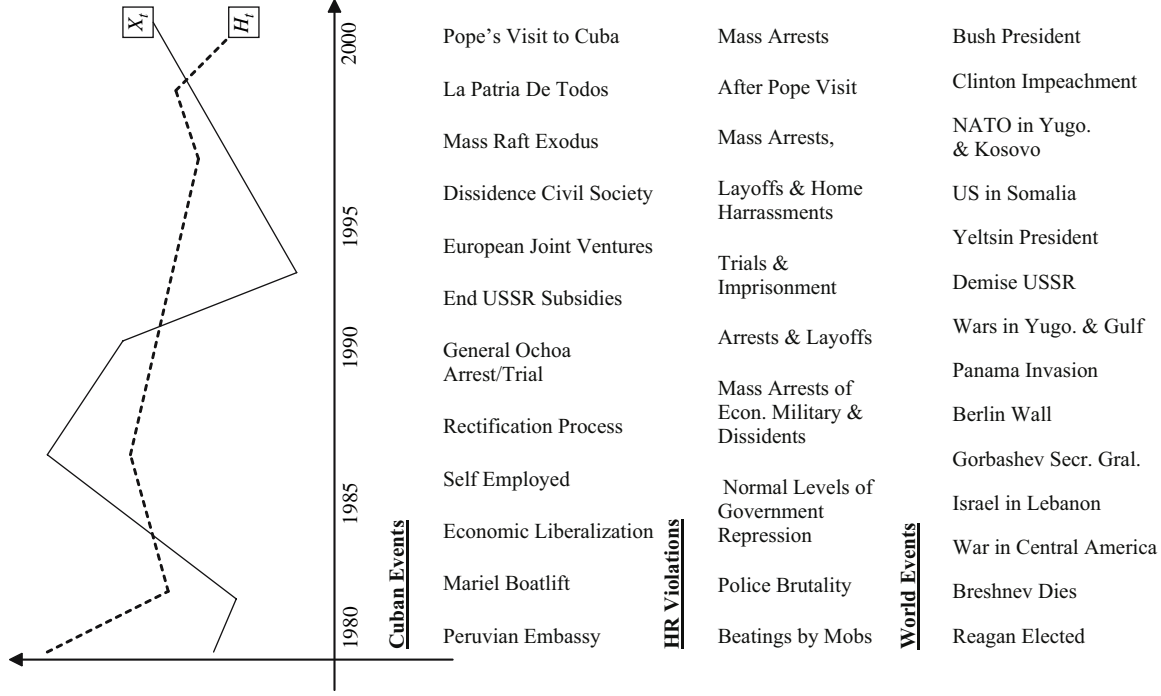


Fig. 4.1 Graphical/analytical representation of HR data, with concomitant (historical events) variables

Note: H_t describes HR violation levels; X_t describes economic development levels

the mid-1960s. All of this triggered another mass movement to leave the country at all costs, which caused extensive riots and the ensuing 1994 Raft Exodus.

Jointly with this social, political, and economic situation, an explosion of peaceful dissent rapidly spread (e.g., publication of the document “La Patria es de Todos,” opening of “free libraries,” organization of independent labor unions and peasant cooperatives, etc.). Those activities culminated with the visit of Pope John Paul II to Cuba, and the rally of one million people in Havana to attend the Pope’s mass. Such political events had not been witnessed in Cuba since the beginning of the revolution.

Throughout that same period (1980–2000), the American Embargo remained in place. Its effect on the overall Cuban economy can be assessed through the model. It appears to be non-significant, as the response is barely affected by it.

To illustrate this modeling approach, we show the Cuban Gross Domestic Product (GDP) by year in Table 4.6. Concomitant events included in the graphical model of Fig. 4.1 are the Peruvian Embassy event, the Mariel Exodus, Free Markets, Rectification Process, the Ochoa trials, internal dissidence, the end of the Soviet Union and its subsidies, European tourism industry, and the Pope’s visit.

Figure 4.1 shows the economic variable X_t , and a second generic HR variable H_t that encompasses incarcerations, detentions without trial, mob attacks on dissidents, etc. Both variables capture the effects of the factors and events of interest. We use them to study their association with, as well as how they are impacted by, the concomitant socioeconomic and historical variables.

A careful look at this time series and its concomitant variables, as well as at the history of the entire 1960–2000 period, shows how, as the economy improves and the people become less dependent on the government, there is an increase in repression (HR violations) followed by a prohibition of the independent economic activities that bring about such independence. This reaction allows the government to regain its control of the political and economic life of the country.

Such cyclic periods of economic freedoms, followed by periods of independence and internal dissent, and then of government crack-downs on the independent economy, have occurred throughout 1980–2000, from the Mariel Boatlift to the Pope’s visit. Such an analysis and modeling approach can help us understand the complexities in the HR situation of a country, and through that understanding devise policies that may eventually lead to finding a solution to the HR problems.

Table 4.6 Example of time series (Cuba) with concomitant variables

Year	1979	1980	1981	1982	1983	1984
X_t	4.2	2.2	16.0	3.9	4.9	7.2
Concomitant events		Mariel Exodus	Free markets			Free markets
Year	1985	1986	1987	1989	1992	1995
X_t	4.6	1.2	-3.6	N/A	N/A	N/A
Concomitant events	Free markets	Rectification process	Glasnost	Ochoa trials	USSR ends	EU tourism

Source: Wilkie (1990, Table 3407)

4.6 Other Statistical Studies and Researchers

HR is a very elusive concept, but a very concrete reality. Like beauty, HR issues are difficult to define but easy to recognize.

Many American Statistical Association (ASA) statisticians have been active, both as interested professionals and as committed researchers, in HR work. Some of them have even been mentors to the author, encouraging him in his work, providing him readings and direction, and co-authoring some of his papers.

The ASA as an institution has also been very active in HR through its Scientific Freedom and Human Rights (SFHR) Committee, to which the author belonged for several years, as well as by promoting good statistical research. Much of this research has been published in the ASA Proceedings of the Social Statistics Section. A Compendium, including papers from 1984 to 2001, is available on the Web: www.amstat.org/sections/ssoc/SSS_Human_Rights_Papers.pdf. The Compendium is a good starting point to read the many excellent papers appearing in all those sources; those papers include interesting case studies and implement many statistical procedures in very unusual ways.

To avoid needless repetition, we do not give in this chapter, which is part of a monograph on the HR subject, references to this work. This in no way diminishes the valuable contributions of all of these researchers to the area of the statistical study of HR violations.

The AAAS has also been in the forefront of HR work. On its Web site there is a HR Directory that includes a search engine and many links to other HR sources (<http://shr.aaas.org/dhr/>). Jana Asher, a co-editor of this book, has also compiled a partial list of research papers on statistics and HR assessments (www.geocities.com/janalynnasher/hrbooklist.html). Asher has also conducted several AAAS in-depth studies about HR violations, where she incorporates some time series charts with analytical and concomitant variable analyses like the ones we have discussed in this chapter (http://shr.aaas.org/pubs/author.php?_id=61).

Without diminishing the work of any others, the author wants to acknowledge those who have mentored him, and have contributed in one form or another to his research. They include D. Banks, T. Jabine, F. Leone, H and L. Spirer, D. Samuelson and J. Asher, among others.

It would not be fair to conclude without mentioning the series of reports by Humana, as well as those of Human Rights Watch, America's Watch, Amnesty International, Freedom House, the US State Department, the United Nations, and other international organizations that the author often uses to obtain data to pursue his work. The HR monitoring program of Columbia University, among others, has produced monographs and reports that are worth examination. The Human Rights Quarterly special issue and the book to which it led (Claude and Jabine 1992) are major milestones in the vast work of statistics in support of HR. Finally, ASA Past-President F. Scheuren has written about the use of survey sampling in HR research. That attests to the degree and comprehensive commitment that the ASA has for HR causes and HR assessment.

4.7 Conclusions

Assessing HR violations is difficult, due to the multi-dimensionality of the problem, the lack of HR data, and measurement scale weaknesses, among other issues. Statistical analysis, with its methodological tools, can be very helpful in uncovering and testing HR violation patterns. In addition, statistical thinking can be useful in interpreting HR data analysis results and in providing an unbiased context in which to conduct a constructive discussion; one where HR violations will not be “justified” as part of the cost of obtaining a “greater” public good.

Summarizing, statistics can contribute to the HR work in at least three important ways. Firstly, by raising *awareness*, among analysts and the public, to different HR factors and their complexities, and to the appropriate use of specific data analysis methodologies. Secondly, by providing a *scientific framework* (statistical thinking and philosophy) where data analyses may be performed in a more unbiased and acceptable fashion. Finally, by incorporating useful *statistical tools* such as case/control methodology, longitudinal studies, and the use of historical and socioeconomic information, jointly with graphical analysis, one can do analyses that take proper account of key concomitant variables. Such inclusion can shed additional light, enlarging the problem dimension and facilitating the search for solutions. The methodologies used here were illustrated via the case of Cuba. However, they may be modified or extended and used in the study of other countries.

Finally, we also discussed the problem of interpreting HR results. The weights α_i , assigned to reduce the problem of dimensionality, are of crucial importance, for different analysts may use different weights in their conscious or unconscious efforts to justify or condemn HR violations. We believe that it is in this arena that statistical thinking can contribute most to HR work.

In his 22 years of direct exposure to HR violations, as well as his 25+ years working in favor of the recognition of HR for all, the author has found one issue of continuous re-occurrence. It is the one related to the “justification” of HR violations by way of pursuing “a greater good.” In the parlance of such HR violators, “violations are an unfortunate necessary evil” in the quest of a “higher, nobler objective.” Such justification pursues two well-defined objectives: one internal and another external. It tries to convince fellow countrymen that certain despicable actions have some moral validity and they should accept this. It also attempts to convince the international community that such HR violations are necessary evils that are forced upon the violator by the actions of foreign governments or by the violator’s desire to raise socioeconomic or political performance measures, at all costs, in a short period of time. The author finds these arguments morally unacceptable.

The author can summarize his HR research work as follows: finding methods for fairly evaluating such HR ambiguities. HR violations are sometimes uncovered and documented, but this that does not necessarily lead to public condemnation for the reasons discussed above and, thus, little action follows. As a result, people suffering HR violations do not improve their situation, which is the main objective of the author’s HR research work. By contributing to take “wind away from the sails” of

supporters of HR violator, whoever those may be, the author is actively contributing toward improving HR causes, for the HR violators will then become isolated.

It is not enough to point out that HR violations occur, or to quantify them. It is also necessary to move others to act. And this is only possible when we can also show that there are no HR violations that can be justified under the “greater good” umbrella.

Acknowledgment Many friends and colleagues have contributed to the author’s research in various ways throughout time. He gratefully acknowledges all of them, both in the Association for the Study of the Cuban Economy (ASCE) and in the ASA. He also thanks the two reviewers for their constructive criticism and comments, and particularly the monograph editors for their careful English editing of an earlier version of this chapter and for their invitation to contribute a chapter to this important book.

References

- Alonso JF, Lago A (1994) A first approximation design of the social safety net of a democratic Cuba. In *Cuba in Transition: Volume 4*. Association for the Study of the Cuban Economy (ASCE), Miami (<http://lanic.utexas.edu/la/cb/cuba/asce/cuba4/healsys1.html>, accessed May 6, 2007)
- Claude RP, Jabine TB (eds) (1992) *Human rights and statistics: getting the record straight*. University of Pennsylvania Press, Philadelphia
- Comité Estatal de Estadísticas (1979) *Anuario Demográfico*. Comité Estatal de Estadísticas, La Habana
- Eberstadt N (1986) Did Fidel fudge the figures? Literacy and health: the Cuban model. *Caribbean Review* 15:4–7, 37–38
- Gordon AM (1983) The nutriture of Cubans: historical perspective and nutritional analysis. *Cuban Studies* 13:1–34
- Humana Ch (1992) *World human rights guide* (3rd edn). Oxford University Press, New York (www.bsos.umd.edu/gvpt/lpbr/subpages/reviews/humana.htm, accessed May 6, 2007)
- Mesa Lago C (ed) (1971) *Revolutionary changes in Cuba*. University of Pittsburgh Press, Pittsburgh
- Mesa Lago C (1981) *The economy of socialist Cuba: a two decade appraisal*. University of New Mexico Press, Albuquerque
- Mesa Lago C (1987) *Cuba in the 1970’s*. University of New Mexico Press, Albuquerque
- Mesa Lago C (2000) *Market, socialist and mixed economies: comparative policy and performance of Chile, Cuba and Costa Rica*. Johns Hopkins Press, Baltimore
- Oficina Nacional de Estadísticas (1971) *Censo de Población y Viviendas*. Comité Estatal de Estadísticas, La Habana
- Oficina Nacional de Estadísticas (1981) *Censo de Población y Viviendas*. Comité Estatal de Estadísticas, La Habana
- Oficina Nacional de los Censos Demográfico y Electoral de Cuba (1954–1955) *Censos de población. Electoral y de viviendas de 1953*. Comité Estatal de Estadísticas, La Habana
- Romeu JL (1993) Statistical thinking can enhance Cuban human rights assessments. *Proceedings of the social statistics section of the American Statistical Association*, Alexandria, pp 118–122
- Romeu JL (1994) A statistical comparison of Cuban socioeconomic development. *Proceedings of the social statistics section of the American Statistical Association*, Alexandria, pp 159–163

- Romeu JL (1996) More on the statistical comparison of Cuban socioeconomic development. Proceedings of 1995 Association for the Study of the Cuban Economy Conference, Miami, pp 293–301
- Romeu JL (2005) On the Juarez Lincoln Marti International Education Project. *American Journal Math & Management Science* 25:1–20. (<http://web.cortland.edu/matresearch>, accessed May 6, 2007)
- Romeu JL, Ciccimaro J, Trinkle J (2004) Measuring cost avoidance in the face of messy data. Proceedings of the reliability and maintainability symposium (RAMS), Los Angeles, pp 157–162
- Romeu JL, Gloss-Soler S (1983) Some measurement problems detected in software data analysis. Proceedings of computer software and applications Conference (COMPSAC), Chicago
- Romeu JL, Samuelson D, Spierer H (1992) An Assessment of the Human Rights Situation in Cuba. Proceedings of the Social Statistics Section of the American Statistical Association, Alexandria, pp 292–297
- The Statesman Yearbook (1929, 1954, 1992) Palgrave Macmillan, Basingstoke, New York (www.statesmansyearbook.com/public/, accessed May 6, 2007)
- United Nations Yearbook (1948, 1952, 1959, 1969, 1971, 1981, 1993) United Nations, New York (www.un.org/Pubs/whatsnew/ycd02.htm, accessed May 6, 2007)
- Wilkie JW (ed) (1990) *Statistical Abstract of Latin America*, Volume 28. UCLA Latin American Center Publications, Los Angeles

Chapter 5

Hidden in Plain Sight: X.X. Burials and the Desaparecidos in the Department of Guatemala, 1977–1986

Clyde Collins Snow, Fredy Armando Peccerelli, José Samuel Susanávar, Alan G. Robinson, and Jose Maria Najera Ochoa

5.1 Introduction

This is a study of two groups of people, *X.X.s* and *desaparecidos*. The members of both groups have been dead for over 20 years. In addition to being dead, the *X.X.s* and the *desaparecidos* have some other things in common. First, they were citizens of Guatemala, where, from about 1960 onward, thousands were killed in nearly four decades of brutal civil conflict. Second, they died sometime between 1977 and 1986 when the violence reached its peak. Finally, some mystery surrounds them all: for the *X.X.s*, the mystery lies in their identity—we don't know *who* they are; for the *desaparecidos*, the mystery is their ultimate fate—we don't know *where* they are. Some readers will also be mystified by the names of these groups. What is an *X.X.*? What is a *desaparecido*?

In Guatemala, unidentified bodies are medicolegally designated by the initials *X.X.* and are hence the equivalent of the “John (or Jane) Does” of the English-speaking world. Such unfortunates are buried at public expense in municipal cemeteries.

During the conflict, many thousands of Guatemalans were killed—some by left-wing guerillas but the majority by agencies of the government. The whereabouts of most are known—they lie in single or mass graves throughout the country and, to date, over 3,000 of their skeletons have been exhumed and examined by forensic anthropologists. But other thousands were abducted by military or police “death squads” and never seen again—they became *desaparecidos* (“disappeared ones”). They differ from those killed outright in that, while it is virtually certain that they are dead, their final resting places are unknown.

In this study, we propose a theory that the *X.X.s* and *desaparecidos* have one more thing in common: overlapping membership or, simply stated, that hundreds of *desaparecidos* lie buried as *X.X.s* in a single place. This place is *La Verbena*, a large municipal cemetery in Guatemala City.

5.1.1 The Cemetery of La Verbena

La Verbena, founded in 1939, is located in a largely working class neighborhood of Guatemala City. It contains few of the elaborate family mausoleums where well-to-do Guatemalans bury their dead. Many of the dead are buried in single graves marked with modest wooden or concrete crosses and hundreds of others are interred in coffin-sized concrete crypts (*nichos*) arranged in multi-storey blocks—like public housing for the dead. On an ordinary day dozens of people come there, some to bury their dead and others, often carrying small bouquets purchased from the flower vendors who line the street outside the graveyard, to visit them. They seldom wander far from the tree-lined walks of the area where their dead are buried.

There is also a large sector in La Verbena seldom visited by anyone but gravediggers. From a distance, it appears to be vacant and untended. It contains few marked graves and is overgrown with the weeds that crop up quickly in the subtropical climate. But a closer inspection of this desolate area reveals many unmarked graves. It is the place where the *X.X.s* are buried. If our theory is true, it is here, that for more than two decades, hundreds of *desaparecidos* have, in effect, been hiding in plain sight.

5.2 Medicolegal Death Investigation in Guatemala

In Guatemala, as in most countries, sudden, unexplained, or suspicious deaths must be medicolegally investigated to determine the identity of the deceased, the time and place of death, and, finally, how the victim died. If he or she is found dead at the scene, the body is sent directly to a judicial morgue. If still alive, the victim is taken to an emergency treatment center from where, after being pronounced dead by an attending physician, the body is transported to the judicial morgue. Since the final destination is the same, it undoubtedly matters little to the decedent whether the trip to the morgue is direct or includes the short detour through an emergency room. However, for us, the distinction is important. This is because, as will be seen below, it offers an important clue as to whether he or she might possibly be a *desaparecido* or a victim of one of the many misfortunes of ordinary life.

5.2.1 Identification

The Guatemalan police have the primary responsibility for the identification of the deceased. It is usually made from the *cedula*, an internal passport which citizens are required to have in possession at all times. It contains the bearer's name, date and place of birth, height, photograph, and fingerprints. When found with the deceased, positive identification can be established by matching fingerprints from the body

with those of the *cedula*. In view of its importance in this regard, when a body is found without a *cedula* there is a strong possibility that it was removed to foil identification.

5.2.2 *Pathological Examination*

At the judicial morgue, the body is examined by an officially appointed physician to determine the *cause* of death. The examination may vary from a simple external inspection when the cause is obvious (e.g., traumatic injuries received in a vehicular accident) to a complete autopsy when it is obscure or complicated. Based on his findings, the medical examiner prepares a report that is forwarded to an investigating magistrate. The latter, after reviewing information supplied by witnesses, the police, and the medical examiner, makes a determination of the *manner* of death, that is, whether it was natural or, if violent, the result of accident, suicide, or homicide.

5.2.3 *Disposal of Unidentified Bodies*

If a body cannot be immediately identified, it is sent to a municipal cemetery where it is buried as an X.X. Interment is prompt since the law requires that unembalmed bodies be buried within 48 hours of death. Cemetery administrators record the exact location of the grave so that if a body is eventually identified by fingerprints or other means, it can be exhumed and turned over to the family.

5.3 Materials and Methods

This study deals with X.X. burials in the Department of Guatemala which encompasses the national capital, Guatemala City, and three smaller municipalities, Mixco, Villa Nueva, and Amatitlán. It is the governmental, economic, and cultural center of the country. Although small (2,126 square kilometers), it is the most densely populated and highly urbanized of Guatemala's 22 departments. According to census records, the 1981 population was 1.786 million.

In the Department of Guatemala, all X.X.s are buried in *La Verbena*. Although the graves are unmarked, records are kept of their location by row and plot number. For each burial, identified or X.X., a notarized burial certificate, the *acta*, is completed. The numerically sequential *actas* are bound in large ledgers indexed by the surname of the deceased and a reference to his or her *acta* number (X.X.s are listed under letter "X"). In August 2004, the authors obtained permission to copy the X.X. information from the *La Verbena* records.

5.3.1 Records Analyzed

This survey covers the period from January 1, 1977 through December 31, 1986. This 10-year period was chosen because it spans the years when the violence of the Guatemalan civil conflict reached its peak.

In all, 3,307 X.X.s were buried in *La Verbena* during the survey period. Data from each burial certificate were entered into a MS Excel database spreadsheet. After preliminary editing, the data were transferred to a MS Access database. Eighty records with incomplete or ambiguous data and those of 56 infants (mostly stillborn) were excluded, leaving 3,171 records for analysis.¹

5.3.2 Variables

The data in this study consists of *primary* and *secondary* variables. Primary variables are entries from the burial certificates. Secondary variables are derived from the primary variables.

The primary variables are as follows:

- Administrative: Burial records are organized chronologically by libro, folio, and acta, the latter being the number of the individual burial certificate.
- Sex: As determined by the medical examiner: male, female, or undetermined.
- Age: Age in years as estimated by the medical examiner.
- Date of death: As given by the medical examiner, n.b. by convention, this is the date that the body is first viewed and pronounced dead by the medical examiner. The actual date of death may have been earlier by days or weeks in the case of decomposed bodies or even years if the remains are skeletonized.
- Date received: Date of arrival at La Verbena.
- Date buried: Date of burial.
- Location: Street address or other specific information denoting the location of the body when it came to the attention of medicolegal authorities. Note, however, that the discovery scene may not be the one where death actually occurred since some victims may have died in one place and had their bodies dumped at another.
- Cause: The cause of death as determined by the medical examiner.
- Notes: Supplemental information entered on the burial certificate.

The secondary variables are as follows:

- Semester: The abstracted records consisted of all X.X. burials from January 1, 1977 to December 31, 1986, a period of 10 years. For analytic purposes, the data are treated in 0.5 year units, yielding 20 *semesters*. For example, semester 77.2 includes all X.X. burials recorded during the last 6 months of 1977.

¹ For the remainder of this chapter, we shall call these data the Guatemala X.X. Study Database (GSD)

- Period: For reasons that will be made clear below, the 20 semesters of this study fall into two periods. The 11-semester *baseline* (BASE) period comprises the first four (77.1–78.2) and last seven semesters (83.2–86.2). It is interrupted by the 9-semester *elevated* (ELEV) period (79.1–83.1).
- Regime: Governmental regime during which death occurred.
- Place pronounced: Whether the victim was pronounced dead at the scene where the body was found (PAS) or at a hospital (PAH).
- COD: From a review of the causes given by the medical examiner, the deaths are classified into the following *secondary* variables:
 - Natural causes (NAT)
 - Violent causes (VIO)
 - Gunshot wounds (GSW)
 - Asphyxiation (ASP)
 - Edged instruments (EDG)
 - Generalized trauma (GNT)
 - nCOD: In cases of violent death, whether the traumata recorded was *single* or *multiple*.
 - MUT: Evidence of postmortem mutilation.
 - UDIK: Unidentified killed listed in the CIIDH database.

5.3.3 Data Limitations

The burial certificate data have two limitations:

1. Cause: The cause of death entered on the burial certificate is the summary diagnosis of the medical examiner and does not necessarily include secondary findings included in his full report. For example, if a body exhibited long-term physical abuse from torture but died of a gunshot wound, the former finding may or may not be found on the burial certificate.
2. Manner: The burial certificates offer no clue to the *manner* of death, that is, whether it was due to homicide, suicide, accident, or undetermined. The difficulty this presents is illustrated by gunshot wounds. In such cases, there is no clue to whether the victim was murdered, committed suicide, shot accidentally, or the evidence was insufficient to make a determination. Presently, the reports listing the manner of death, scattered in magisterial archives throughout the Department, are not readily accessible.

5.3.4 Assumption of the Present Study

In modern societies the number of unidentified bodies that turn up annually is a non-random subset of overall mortality. Thus, factors influencing the total number of people of a city, province, or country who die each year also act on the component

of these deaths that go to their graves unidentified. The absolute number of unidentified dead varies widely among societies. The most obvious factor influencing that number is population size: all things being equal, more people die each year in large populations than in small and, hence, larger numbers of the dead are apt to be unidentified. Also important are conditions affecting general health and the sophistication of the medicolegal system charged with the task of identification of the victims of violent death. As a result, the annual number of unidentified dead may be many times higher in poor countries like Guatemala than in healthier and wealthier nations.²

An underlying assumption of the present study is that, within a given jurisdiction, the normal “load” of unidentified bodies tends to remain fairly constant from year to year and that a significant departure from this baseline suggests an exceptional agent at work. Such agents might include natural events like hurricanes, earthquakes, floods, and epidemics that overwhelm the capacity of the local medicolegal death investigation system to deal with mass casualties. But not all disasters are natural: man-made events such as serial or mass murders, warfare, and political violence can produce similar results. The isolation and identification of such factors can be approached in much the same manner as epidemiologists isolate and identify the etiological agents of a natural disease.

As examples of the relative stability of the normal load of unidentified bodies, we have calculated statistics over a 10-year period from four highly urbanized jurisdictions comparable in size to the Department of Guatemala (Table 5.1).³

The third column in Table 5.1 shows average annual number of unidentified bodies as a function of population size in each jurisdiction. Its almost sixfold range represents the interaction of many demographic (age, sex, ethnicity, etc.) and social (poverty levels, crime rates, etc.) variables as well as the overall effectiveness of the jurisdiction’s death investigation system. Despite the foregoing differences, the number of cases observed from year to year is relatively low, ranging from zero to fourteen. This variation represents chance fluctuation from the baseline. Suppose, however, that a catastrophic event such as a hurricane or terrorist bombing resulted

Table 5.1 Unidentified bodies found in four urbanized jurisdiction in the United States and Europe, 1995–2004

Jurisdiction (and major city)	Population (millions)	Unidentified per million	Annual cases	Min	Max
Clark County, Nevada (Las Vegas)	1.38	5.96	8.2	2	14
Fulton County, Georgia (Atlanta)	0.82	3.68	3.0	0	5
Harris County, Texas (Houston)	2.82	1.03	2.9	0	12
Province of Milan (Milan)	3.71	1.78	6.6	3	10

Source: www.doenetwork.org/uidlinks.html (accessed May 6, 2007).

² In 1996, the homicide rate in Guatemala City was 101.5 per 100,000 inhabitants, placing it third among major urban areas of the Western Hemisphere after the Columbian cities of Medellin and Cali with 248 and 112 per 100,000, respectively (Piquet Carniero 2000).

³ Compiled from data provided through <http://www.doenetwork.us/uidlinks.html>.

in massive casualties among which many could not be identified. This would produce a strong increase in the number of unidentified bodies for the year in which the event occurred.

To illustrate, we have constructed a hypothetical jurisdiction based on the pattern observed in the four jurisdictions shown in Table 5.1. In this example, the number of unidentified bodies reported normally fluctuates randomly around an average of about five per year. But in the fourth year, 45 unidentified bodies are reported. As shown in Fig 5.1, this would produce in a sharp departure from the normal baseline. Clearly, such a transient peak signals a highly unusual event during the year in question, and that finding would trigger a search for the etiological agent responsible. Closer investigation of the records of year 4 would reveal the event responsible—perhaps a mass disaster such a hurricane, or even the activities of an unusually prolific serial killer.⁴

In the example in Fig. 5.1, we have set the normal load of unidentified bodies to five—which appears to be fairly typical of the jurisdictions shown in Table 5.1 characterized by low crime rates and sophisticated medicolegal death investigation systems. In jurisdictions where those conditions do not apply, the baseline might be much higher but one would still expect to find a pattern of relatively low random fluctuations around the mean in normal years. Significant departures from the normal regression would excite the curiosity of an epidemiologist.

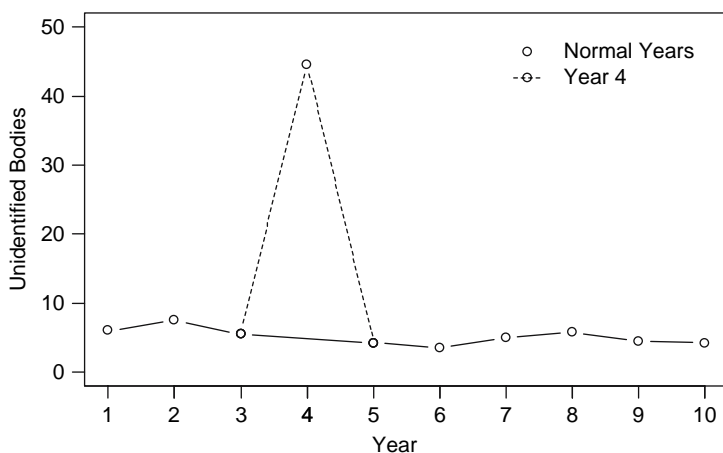


Fig. 5.1 Hypothetical jurisdiction
Source: Simulated data

⁴ Perhaps Hurricane Katrina, which hit New Orleans in August 2005, provides another example. Although normal baseline values are lacking for New Orleans and surrounding jurisdictions, it is safe to presume that they are not too different from the urbanized jurisdictions tabulated above. Five months after the storm, 114 of the 1,417 officially reported victims had not been identified (New York Times, February 2, 2006).

5.4 Results

The distribution of X.X. deaths by semester is shown in Fig 5.2. Its overall pattern shows two upward fluctuations. The first, and longest, began in January 1979 and ended in July 1983. It is bimodal with a peak of 217 deaths in early 1979 and another of 358 in the latter half of 1981. The second, more modest, increase occurred in 1986 when deaths increased to about 160 per semester. These upsurges rise from a baseline of slightly over 100 deaths per semester.

We will assume that the baseline represents the number of X.X. deaths normally expected and, conversely, significant departures are anomalous. Linear regression analysis of the baseline values yields the following equation where D_p is the number of deaths predicted:

$$D_p = -1.152 (\text{fractional year}) + 2389$$

Thus, during semester 81.2, the predicted number of X.X.s would be

$$D_p = -1.152(1981.75) + 2389$$

$$D_p = 105.2$$

Yet, in fact, 358 X.X.s—more than three times the number predicted—occurred during semester 1981.2. This number far exceeds the upper 99 percent prediction limit for the equation. Subtracting the predicted number of deaths from those actually observed (D_o) leaves the number in excess (D_x) of those predicted:

$$D_x = D_o - D_p$$

$$D_x = 358 - 105$$

$$D_x = 253$$

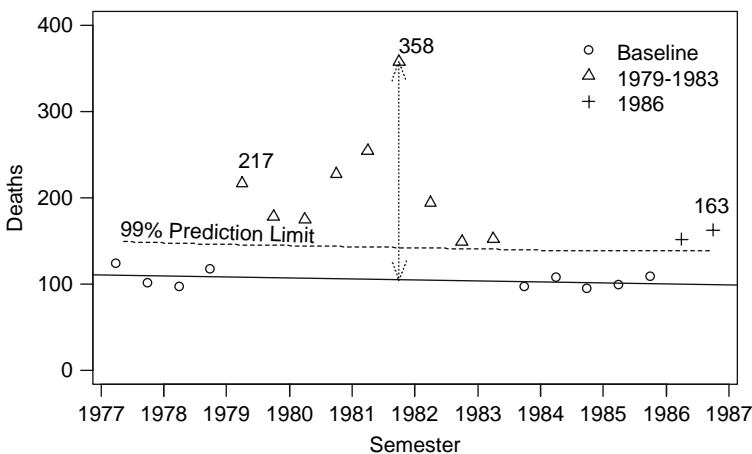


Fig. 5.2 All X.X. deaths by semester (N=3,171)
Source: GSD

In other words, during the last 6 months of 1981, 253 more X.X.s were buried than would be normally expected. Altogether, the estimated number of excess X.X.deaths is 1,054—one-third of the 3,171 recorded. Who were they? How did they die? These questions are explored below.

5.4.1 Natural Deaths

The 806 deaths from natural causes make up close to one-fourth of the series. Diagnoses range from infectious diseases such as tuberculosis to chronic conditions associated with malnutrition, alcoholism, and degenerative diseases of old age. Those dying of natural deaths were older than those who died violently by more than a decade and about 87 percent were male. The age and sex differences between the two groups are statistically significant (Table 5.2).

Natural deaths averaged 40.3 ± 13.8 per semester, remaining constant until the last semester of 1986 when they rose above the 95 percent prediction limit. This increase accounts for the rise in the overall number of X.X. deaths during the same year (Fig 5.3).⁵ It is clear, however, that they did not contribute to the major increase between 1979 and 1983. In other words, this earlier rise must have been due to the other major component of our series: violent deaths.

5.4.2 Violent Deaths

A total of 2,365 X.X.s died violently. They can be divided into two groups depending on whether the victim was pronounced dead at the scene (VPAS) or at a hospital (VPAH). The latter numbered 520 or 22 percent of all violent deaths. They

Table 5.2 Natural and violent X.X. deaths by sex and age

Variable	Natural	Violent	Total	Statistical tests	df	P
Mean	46.05	32.50		Unpaired t (Welch)		
Age S.D.	13.97	11.88		t = 24.668		
N	806	2,365	3,171		1, 224	<0.0001
Sex Male	700	2,132	2,832	Chi-Square (Yate's)		
Female	106	233	339	$\chi^2 = 6.513$	1	<0.0107

Source: GSD

⁵ While it is beyond the scope of this study to determine the factor(s) responsible for this 1986 increase in natural deaths, one might speculate it was at least partly due to the many refugees displaced from the countryside as a result of the massive military campaign conducted in the Highlands and other remote areas under the Rios Montt regime in 1982–1983. Targeted primarily against the Mayan indigenes, it resulted in the destruction of hundreds of villages. While thousands of survivors fled to Mexico, many found refuge in Guatemala City. Destitute, malnourished and often alone, many, no doubt, succumbed from poverty-related natural causes in the streets and alleys of the capital. With no one to identify their bodies, such unfortunates were highly likely to be buried as X.X.s.

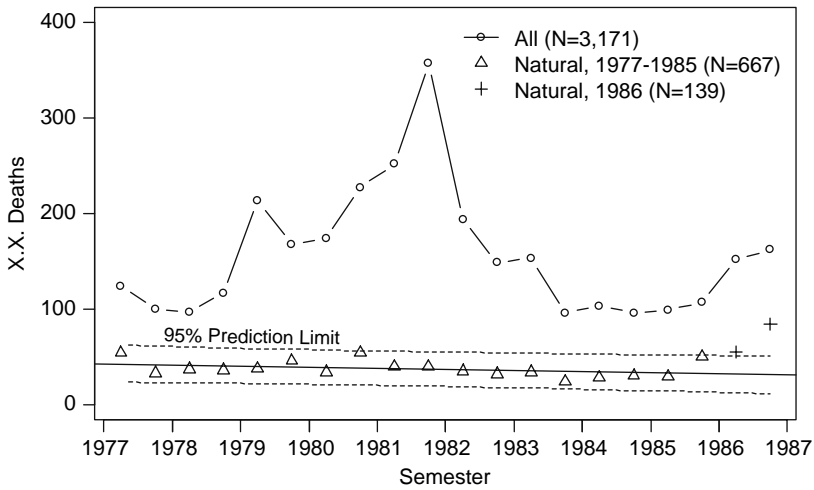


Fig. 5.3 Natural and all X.X. deaths, 1977–1986
Source: GSD

remained low and declined slightly during the survey period, averaging 26 ± 7.7 per semester. As shown in Figure 5.4, they did not contribute to the 1979–1983 rise in X.X. deaths.

One thousand eight hundred and forty-five victims were pronounced dead at the scene, outnumbering VPAH deaths by more than three to one. In contrast to the VPAH deaths, the temporal distribution of the VPAS series is highly correlated with the overall distribution of X.X. deaths ($r = 0.97, r^2 = 0.94, p < 0.0001$). From this it

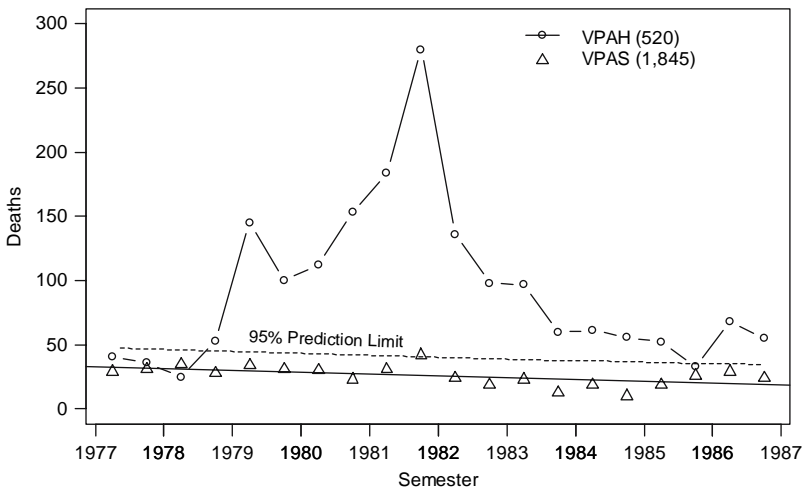


Fig. 5.4 Violent deaths pronounced at scene (VPAS)
Source: GSD

Table 5.3 Distribution of VPAS deaths by age and sex

Cause of death	Count	Percent	Age		Sex		Percent female
			Mean	S.D.	Male	Female	
Gunshot wound	673	36.5%	28.18	7.65	627	46	6.8%
Asphyxiation	387	21.0%	30.33	9.10	340	47	12.1%
Edged instrument	259	14.0%	28.85	9.09	243	16	6.2%
Generalized trauma	526	28.5%	35.47	12.89	469	57	10.8%
Total	1,845	100.0%	30.80	10.34	1,679	166	9.0%

Gunshot wounds (GSW's): Two deaths were caused by shotguns (escopeta); the rest were bullet wounds. Burial records do not indicate whether bullet wounds were caused by handguns, machine pistols, or rifles.

Asphyxiation (ASP): The 387 asphyxial deaths include 225 strangulations (manual and ligature), 86 hangings, 56 drownings, 11 from suffocation and 9 from aspiration of foreign bodies. The higher incidence in females is statistically significant (Fisher's exact p (2-sided) = 0.021).

Deaths from edged instruments (EDG): These 259 deaths include lacerations and stab wounds from weapons ranging from knives to machetes.

Generalized trauma (GNT): Five hundred and twenty-six VPAS deaths were due to generalized trauma. In these, the examining physician simply described contusions, lacerations, and/or fractures without giving any indication of how they occurred. In general, they are typical of injuries sustained in vehicular, industrial, and domestic accidents. Undoubtedly, they also include blunt force trauma homicidally inflicted in altercations such as fistfights and in the course of spousal or child abuse. Also, some may have occurred while the victim was incarcerated.

GNT victims were significantly older than those of other forms of VPAS death (ANOVA, 1-way, $p < 0.0001$) and their ages varied more widely (Table 5.3). This age difference probably reflects the fact that many and, perhaps the majority, of GNT victims died from falls, pedestrian/vehicle impacts, and other mishaps to which the very young and the elderly are particularly susceptible.

Source: GSD

is obvious that the upsurge in X.X.s buried in *La Verbena* between 1979 and 1983 was due almost entirely to VPAS victims.

The distribution of VPAS deaths by age and sex is shown in Table 5.3 The 673 gunshot wounds are the largest component of the series (36.5 percent), followed by deaths from generalized trauma (28.5 percent), asphyxiation (21 percent), and edged instruments (14 percent).

The categories of VPAS deaths by semester are shown in Table 5.4. Although differing in pattern and scale, all show values above the baseline for the nine-semester

Table 5.4 Means and standard deviations of VPAS deaths by period

Period	Count	Gunshot wounds		Asphyxiation		Edged instruments		Generalized trauma	
		Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
BASE	11	14.0	4.8	8.7	5.7	6.5	3.3	19.9	6.6
ELEV	9	57.7	35.3	32.3	13.2	20.9	14.9	34.1	9.6
Unpaired t test (Welch corr.) $p = <0.01$				<0.001		<0.05		<0.01	

Source: GSD

period (shown in red) from 79.1 to 83.1. Values exceeding the 95 percent prediction limits of their baseline linear regressions vary: all nine semesters were significantly higher for gunshot wounds, seven for asphyxial deaths, four for edged instruments, and three for generalized trauma. In all groups, the differences between the semester means of the baseline (BASE) and elevated (ELEV) periods were statistically significant (Fig 5.5).

The excess numbers of VPAS deaths during the ELEV period calculated from the baseline regressions for each cause of death are shown in Table 5.5. Nearly half (45.3 percent) were due to gunshot wounds. One-quarter (25 percent) were deaths by asphyxiation; deaths from edged instruments (15.1 percent) and blunt force trauma (14.6 percent) make up the rest. The total from all causes yields an estimate 889 excess *X.X.* deaths during the ELEV period.

5.4.2.1 Gunshot Wounds to the Head

During the BASE period, of 154 VPAS firearm deaths reported, 74 percent exhibited one or more gunshot wounds to the head. This rose to 84.8 percent during the ELEV period. This difference is significant at the 0.01 probability level (Table 5.6). The implications of this finding are discussed more fully below.

5.4.2.2 Single and Multiple Traumata

Table 5.7 shows the incidence of cases exhibiting single or multiple traumata. The latter includes cases involving more than wound of the same type (e.g., multiple stab wounds of the thorax) or more than one type of violence (e.g., GSW *and* GNT). Multiple wounds occur more frequently in homicides than in suicides or accidents.

Multiple traumata due to gunshot wounds, asphyxiation, and edged instruments increased significantly during the ELEV period. This finding indicates a rise in the intensity of the violence in these types of death and is associated with the phenomenon of *overkill*—the infliction of trauma beyond what is required to cause death.

5.4.2.3 Postmortem Alteration of Remains

Thirty-two bodies were altered postmortem, apparently to hinder identification. All but two of the cases—both 1985 decapitations—were reported during the ELEV period. The difference is significant (*Fisher's exact test, p = 0.004*).

Twelve were burned beyond recognition; seven had died of gunshot wounds and five by ligature strangulation. The other twenty cases exhibited postmortem mutilation. In three, the hands had been amputated after ligature strangulation. They were found together outside Guatemala City on the Palencia highway on January 6, 1982.

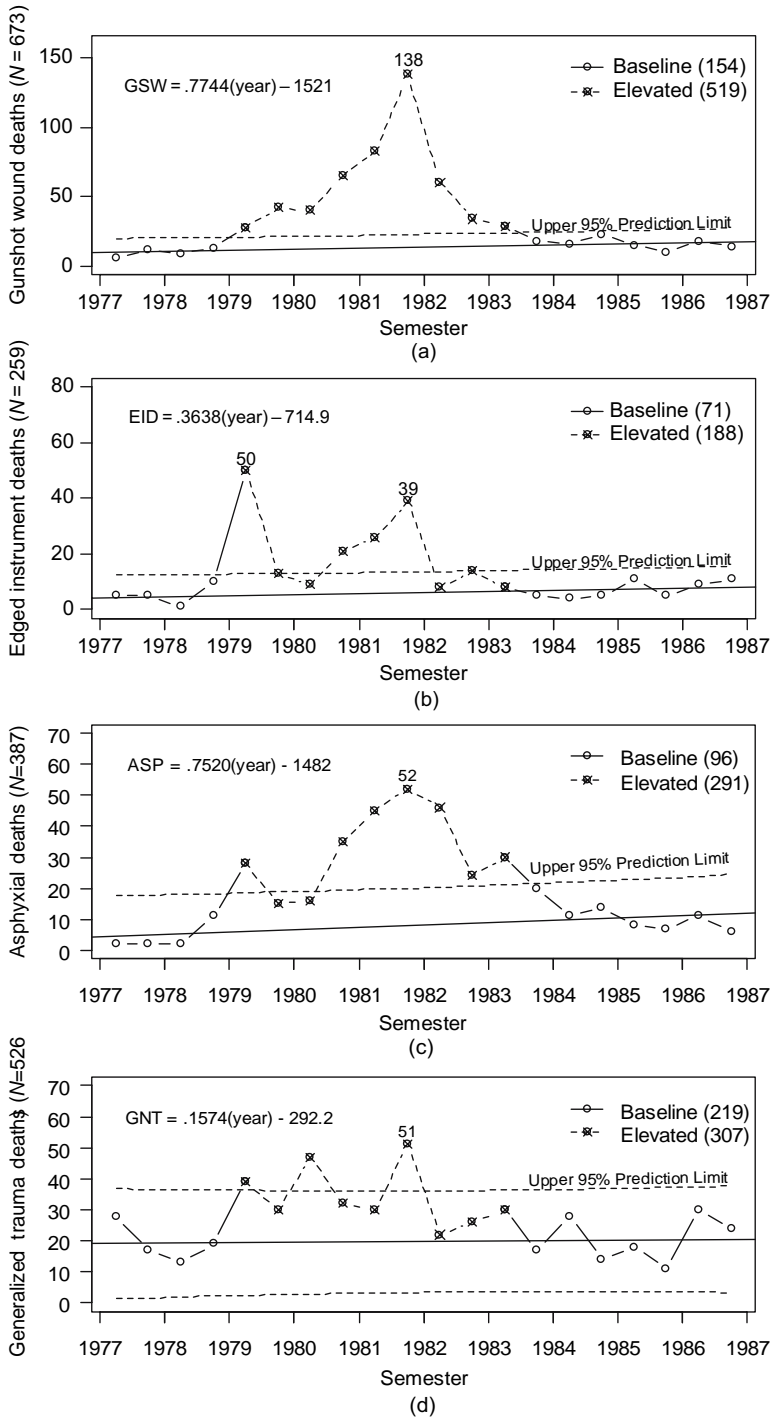


Fig. 5.5 X.X. Burials by cause of death
 Source: GSD

Table 5.5 Excess VPAS deaths during the ELEV period by cause⁸

Semester	Gunshot wounds			Asphyxiation			Edged instruments			Generalized trauma			Total excess
	Predicted	Observed	Excess	Predicted	Observed	Excess	Predicted	Observed	Excess	Predicted	Observed	Excess	
<i>1979</i>													
Sem 1	11.4	28	17	6.1	28	22	5.2	50	45	19.4	39	20	103
Sem 2	11.8	42	30	6.5	15	8	5.4	13	8	19.5	30	11	57
<i>1980</i>													
Sem 1	12.2	40	28	6.9	16	9	5.6	9	3	19.5	47	27	68
Sem 2	12.6	65	52	7.3	35	28	5.8	21	15	19.6	32	12	108
<i>1981</i>													
Sem 1	12.9	83	70	7.7	45	37	6.0	26	20	19.7	30	10	138
Sem 2	13.3	138	125	8.1	52	44	6.1	39	33	19.8	51	31	233
<i>1982</i>													
Sem 1	13.7	60	46	8.4	46	38	6.3	8	2	19.9	22	2	88
Sem 2	14.1	34	20	8.8	24	15	6.5	14	7	19.9	26	6	49
<i>1983</i>													
Sem 1	14.5	29	15	9.2	30	21	6.7	8	1	20.0	30	10	47
Total	116	519	403	69	291	222	54	188	134	177	307	130	889
% Excess	45.3			25.0			15.1			14.6			100

Source: GSD.

Table 5.6 Distribution of VPAS gunshot wounds to head and other body regions by period

Period	Head	Other	Total	Percent head	Fisher's exact test
BASE	114	40	154	74.0	p (2-sided) = 0.004
ELEV	440	79	519	84.8	
Total	554	119	673	82.3	

Source: GSD.

The remaining 17 were decapitated. Six, all males, were killed on August 23, 1981 and found together along the El Salvador highway about 25 km outside the City. Five weeks later, two females in their early twenties were found decapitated alongside the *Ruta al Pacífico* highway 22 km from the capital.

In contrast to the VPAS series as a whole, in which 166 (9.0 percent) of the victims were female, seven (21.9 percent) of these 32 victims were female—a statistically significant difference (*Fisher's exact test*, $p = 0.021$).

5.5 Discussion

As shown above, the increase in X.X. deaths during the ELEV period was due to an upsurge in VPAS deaths. The higher number of deaths displaying multiple traumata, gunshot wounds to the head and instances of postmortem mutilation indicate an increase in the intensity of violence during this 4-year period. Possible reasons for these findings are discussed below.

5.5.1 Ordinary Criminal Activity

During the ELEV period, the rate of VPAS deaths (23.6/month) was almost three-fold than observed during BASE period (8.3/month). At the same time, the number of VPAH victims remained constant. To explain these findings as the result of common crime, we would have to accept the unlikely proposition that, beginning in late 1979, ordinary murderers not only tripled their activities but took extraordinary pains to make sure their victims died at the scene and then, after 4 years of unmitigated mayhem, reverted to their still-lethal but more benign old habits.

5.5.2 War and Terrorism

Civil and military conflict inevitably leaves many unidentified dead. Our survey spans the years when the struggle between the guerrillas and government peaked in violence. Of the 28,513 documented killings and disappearances in Guatemala

Table 5.7 Single and multiple traumata by cause of death

Period	Gunshot wounds			Asphyxiation			Edged instruments			Generalized trauma		
	Single	Multiple	Percent multiple	Single	Multiple	Percent multiple	Single	Multiple	Percent multiple	Single	Multiple	Percent multiple
BASE	88	66	42.9	92	4	4.2	36	35	49.3	132	87	39.7
ELEV	185	334	64.4	246	45	15.5	55	133	70.7	178	129	42.0
Fisher's Exact, $p=$	<0.0001			0.0025			0.0020			n.s.		

Source: GSD

between 1977 and 1986, the names of only 5,714 (20.1 percent) of the victims are known. It is thus logical to explore the possibility that at least some of the X.X.s buried in *La Verbena* were victims of the conflict.

Five governments ruled Guatemala during the survey period (Table 5.8) which includes parts of the Laugerud Garcia and Cerezo Arévalo regimes and entirely encompasses those of Generals Lucas Garcia, Rios Montt, and Mejía Víctores. Three of the five incumbents were elected and two took power by military coup.

Figure 5.6 shows the mean number of VPAS deaths per month during the tenure of each incumbent. Under the rule of General Lucas Garcia, the monthly average of VPAS deaths was 25.2, a nearly fivefold increase over those of his civilian predecessor, Laugerud Garcia. From this peak, they declined to 14.9 per month during the incumbency of General Rios Montt and then to about 10 per month under General Mejía Víctores and Cerezo Arévalo. It is clear that the increase in VPAS deaths during the ELEV period occurred during the regimes of Generals Lucas Garcia and Rios Montt.

5.5.3 State Terrorism in Guatemala

Political violence pitted an ultra-rightist state against extremists of the left. Both sides committed terrorist acts such as assassinations, massacres, and disappearances. However, about 90 percent of the cases so far documented were committed by agencies of the State—military, police, and covert paramilitary groups. The State's war against subversion was conducted in two overlapping campaigns best characterized as *rural* and *urban*. Although the rural campaign, waged primarily against the indigenes of the rugged and remote western highlands, claimed the vast majority of victims, our focus here is limited to the urban campaign which reached its peak in 1980–1982 during the regime of General Lucas Garcia.

Table 5.8 Guatemalan presidential regimes during present study

Regime	Background of incumbent	Entered office by	Dates in office during present study	Tenure during study (years)
Kejell Eugenio Laugerud Garcia (LG1)*	Military	Election	1/1/1977– 7/1/1978	1.50
General Romeo Lucas Garcia (LG2)	Military	Election	7/2/1978– 3/23/1982	3.73
General José Efraín Rios Montt (RIO)	Military	Military coup	3/24/1982– 8/8/1983	1.38
General Oscar Mejía Víctores (MEJ)	Military	Military coup	8/9/1983– 1/15/1986	2.44
Marco Vinicio Cerezo Arévalo (CER)**	Civilian	Election	1/16/1986– 12/31/1986	0.96

*Entered office on 7/2/1974

**Left office on 1/14/1991

Source: GSD.

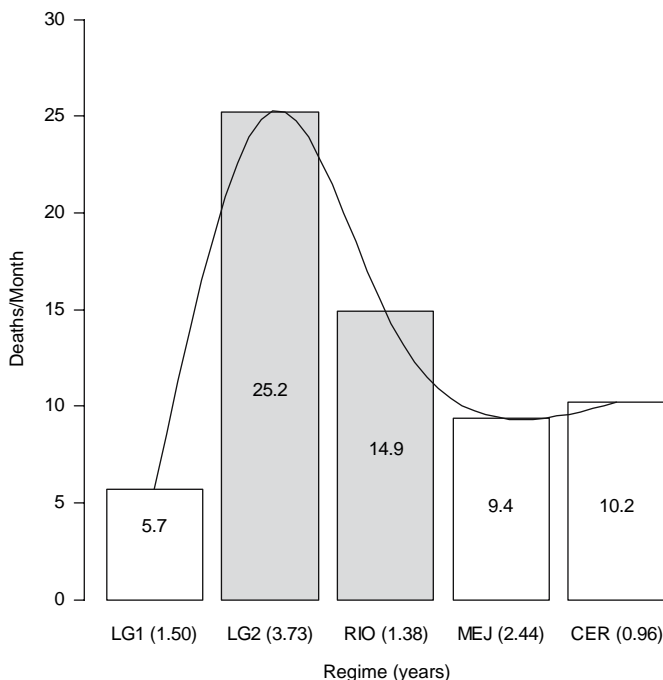


Fig. 5.6 VPAS deaths by register
Source: GSD

The strategy and tactics of the urban campaign were similar to those pursued by the Argentine Junta during its self-proclaimed “dirty war” against subversives, both real and perceived, during the 1970s. Since Argentine anti-subversive operatives came to Guatemala as advisors to the Lucas Garcia government as early as 1980 when the campaign began in earnest, the similarity is not likely to be coincidental (Feldman 1985).

The center of gravity of the urban campaign was the Department of Guatemala. Targeting not only guerillas but members of non-violent dissident movements, it took a heavy toll among intellectuals—lawyers, journalists, and other professionals. For example, although it was a government institution, the University of San Carlos (USAC), Guatemala’s center of higher learning, was particularly hard-hit: from 1977 through 1986, 250 students, 89 professors, and 34 administrative employees were killed or disappeared (Kobrak, P. 1999).

Acts of State violence took several forms. Many were straightforward assassinations; in these, the selected target was simply gunned down at home, on the street, or at the place of work. In others, the victims were kidnapped and within a day or two their bodies—often showing signs of torture—were openly dumped in public places. For instance, many of the USAC victims, both students and faculty, were thrown from speeding automobiles in front of the main USAC administrative building. In such cases, the victims were often well-known critics of the government whose deaths were meant to give dramatic public warning that like behavior would

bring a similar fate. Since the bodies could be promptly claimed and buried by their families, these victims were not interred as X.X.s. In contrast, two other forms of violence, *armed confrontation* and *abduction*, may have resulted in many victims being buried as X.X.s in *La Verbena*.

Throughout the urban campaign, police or military security forces clashed frequently with guerrillas. Such confrontations took place during guerrilla attacks against government targets or during raids by security units against suspected guerrilla hideouts. Most guerrilla units operated as small, independent cells. Such groups avoided committing details of their membership to paper. Even under torture, survivors, if any, could not often identify dead comrades because most knew each other only by pseudonymous “war names.” Thus, when a cell was wiped out, the identity of its members was also lost. In any case, security forces would find it advantageous to conceal the identities of the victims. Consequently, guerrilla deaths were likely to be interred as X.X.s.

Abductions targeted individuals with known or suspected ties to the guerrillas in order to collect intelligence useful in counter-insurgency operations. They were carried out by specialized teams trained in techniques of clandestine kidnapping. The objective was to snare and subdue the victim quickly and silently. The normal routine of the selected target was established through several days or weeks of surveillance in order to determine the optimum time and place of abduction. The kidnapping was carried out by agents in civilian clothes driving unmarked vehicles. Success ensured the victim would be delivered alive to the detention center. Failures could be dangerous: if armed, the target might wound or kill his abductors; if he escaped, he could warn others in his circle that they were jeopardized.

Upon arrival at the detention center, abductees were interrogated under severe torture. The ultimate outcome varied. About 14 percent were “turned” for use as informants or for propaganda purposes. But, for most, their fate was sealed. Release was not an option because, once freed, they could negate any intelligence of value they had disclosed. Instead, they were killed. These victims, unlike those where the motive was to send a “message”, could not be openly dumped in public places. To do so would signal their comrades still at large that they might be compromised by information divulged by the victim.

It was thus important to conceal the bodies. A few came to light weeks or months later in clandestine graves. For example, in early 1980 a gunfight broke out between police and marchers in a funeral procession for victims of the Spanish Embassy massacre.⁶ Police detained several mourners, including USAC student Liliana Negreros. Six weeks later, Liliana’s body was found among 17 others in a mass grave near San Juan Comalapa.

But another way to “conceal” the body would be to strip it of any means of identification and deposit it quietly in some out of the way place in or near Guatemala City. When found, it would be sent to the judicial morgue and thence to *La Verbena* as an

⁶ On January 31, 1980, police stormed the embassy after it had been taken over by armed protesters. In the ensuing gun battle, a fire broke out and police blocked firefighters from entering. Thirty-seven people died, including the embassy personnel held hostage, as well as all of the hostage-takers.

X.X. If this method of disposal were followed in Guatemala, it might account for the fact that after nearly 30 years, the bodies of hundreds of urban *desaparecidos* have never been found. It is interesting that a similar pattern was followed in Argentina where close to 80 percent of the *desaparecidos* have been found in unmarked graves in municipal cemeteries.

In short, we hypothesize that there were two paths followed by *desaparecidos* to their X.X graves in *La Verbena*: first, as victims killed in armed confrontations or, second, as extra-judicially executed abductees (Fig 5.7). Some data from the X.X. series supports this hypothesis.

Normally, in armed confrontations where firearms are used and the rules of conventional warfare are followed, only 20 percent of bullet wounds prove instantly fatal. This is because the most vital parts of the human body such as the heart, great vessels, and brain present relatively small targets. For example, the head constitutes only about 12 percent of the target area of a human body. Combatants wounded in other body regions, although they may die later, can be expected to survive long enough to receive medical treatment if it is promptly available. So constant is this one-in-five ratio of killed-in-action (KIA) to wounded-in-action (WIA) that when it is significantly exceeded, one may suspect the rules of warfare were violated by “finishing off” the wounded. Most commonly, this is done by gunshot wounds to the head. That this practice was followed by Guatemalan counter-insurgency units is indicated by the statistically significant increase in gunshot wounds to the head among VPAS X.X.s during the ELEV period (Table 5.6).

Two decades ago, most firearm wounds in suicides and ordinary criminal homicides in Guatemala were inflicted by handguns—usually revolvers. Automatic rapid-fire weapons were standard weapons of the military. They were also issued

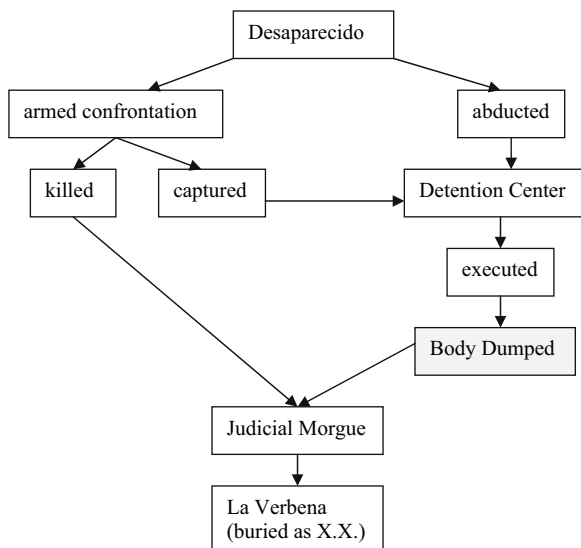


Fig. 5.7 Flowchart depicting possible disposal of *desaparecidos* as X.X.s

to special anti-subversive squads of the police. Due to their high rate of fire, they can deliver many projectiles into the target in a matter of seconds. Their use in armed confrontations with guerrillas may account for the significant increase in multiple gunshot wounds observed in VPAS X.X.s during the ELEV period (Table 5.7).

5.5.4 The Beginning and End of the Urban Campaign

In our analysis, we defined the ELEV period on the basis of whether or not the number of VPAS burials that occurred during a given semester exceeded the 95 percent prediction limit of the regression line of the BASE period. Nine semesters met this strict criterion. Thus, the ELEV period began in 79.1, the second semester of the Lucas Garcia regime, and ended with 83.2, the last full semester that Rios Montt held power. It is unlikely, however, that the actual beginning and end of the urban campaign were precisely defined by these semesters.

Lucas Garcia took office on July 1, 1978. During the first 4 months of his tenure, X.X. burials remained low—averaging 6.5 per month. Then, in November, they more than doubled to 16 and continued to rise thereafter (Fig 5.8). Thus, it appears likely that the urban campaign began in earnest around this time.

The 4-month delay in putting the campaign in action was no doubt related to the fact that such an extensive effort required elaborate preparation. Key to its planning were the nearly 50 Argentine anti-subversive operatives who arrived in Guatemala as advisors a few weeks after Lucas Garcia took office (Feldman 1985). Fresh from their victory in their own self-styled “Dirty War” against the largely urban-base

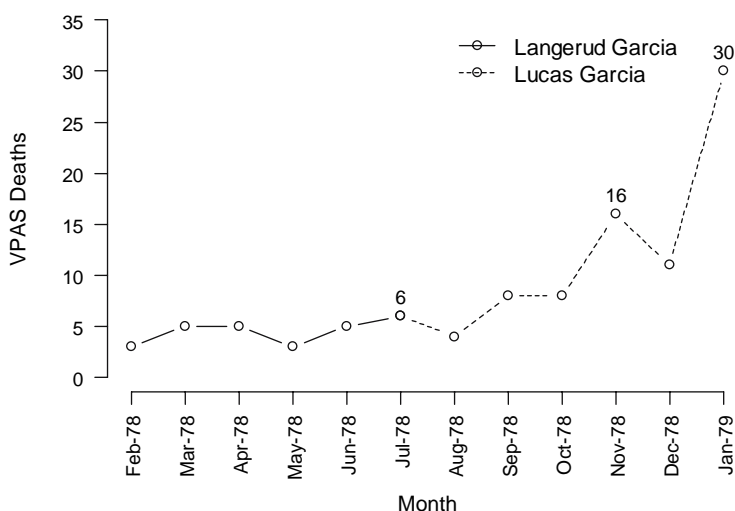


Fig. 5.8 Regime change and initiation of the urban campaign in november 1978
Source: GSD

Montoneros, they apparently played an important role in the Guatemalan urban campaign and, in many ways, the strategy and tactics of Argentine and Guatemalan operations were eerily similar.

General Mejía Victores overthrew Rios Montt in a coup d'état on August 8, 1983. Since he had served Rios Montt as defense minister, Mejía Victores bore heavy responsibility for the excesses committed by the military during the 17-month rural campaign. Once in power, however, he formed a predominately civilian government focused on institutional reform. Disappearances continued but at a much lower rate than during the regimes of Lucas Garcia and Rios Montt.

Perhaps the best indicator of the end of the most intense period of the urban campaign is provided by the monthly numbers of VPAS X.X. burials in La Verbena during the transition period between the two regimes (Fig 5.9). In the last 6 months of Rios Montt's rule, VPAS burials averaged 12.2 per month. When Mejía Victores took over, only two VPAS deaths occurred during the first month of his regime. This abrupt drop probably indicates that executions were put on hold while some changes in urban campaign operations were being considered. After this pause, they rose to 22 in September and 21 in October. In November, they again declined to an average of about 10 per month which was maintained throughout the remaining 2 years of his rule.

The steep rise in VPAS deaths during September and October may signify a "house cleaning" during the closeout of certain detention centers similar to those which occurred in Argentina in 1978 and in 1983 (Snow and Bihurriet 1983). The first was undertaken as an attempt to polish the Junta's tarnished international image before Argentina hosted the 1978 World Cup in soccer and the second when it became clear to the Junta leaders that they were about to lose power after their disastrous defeat in the Falklands War. In both instances, prisoners who had until

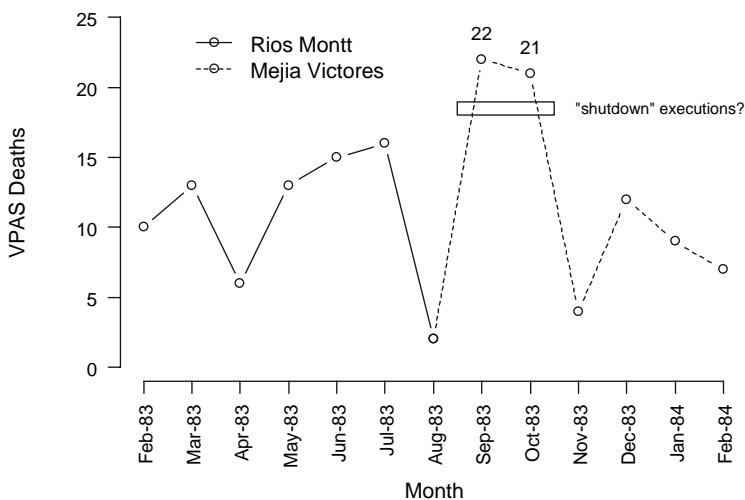


Fig. 5.9 Regime change and end of the urban campaign
Source: GSD

then been spared were finally executed—simply liberating them would prove too embarrassing for the government which had, over the years, officially and loudly denied ever having had them in custody.⁷ It seems likely that the transient rise in VPAS deaths in Guatemala during the early months of the Mejía Victores regime also reflects the execution of prisoners still held in detention centers scheduled to be shutdown as the urban campaign was brought to a close.

5.5.5 X.X. *Burials and Unidentified Victims*

Beginning in 1994, the International Center for Human Rights Investigations (CIIDH) began a systematic compilation of human rights violations that occurred in Guatemala during the 36-year period (1960–1996) of civil conflict. This effort was headed by Patrick Ball, then-Deputy Director of the Science and Human Rights Program of the American Association for the Advancement of Science (AAAS). Building on data previously collected by Paul Yamauchi (1993) and 4,000 archival reports in the files of various Guatemalan human rights groups, Ball and his colleagues documented over 10,000 additional cases of killings and disappearances, most reported in the Guatemalan press. Since many cases—particularly the village massacres in the highlands—resulted in the death or disappearance of more than one person, the total number of victims was 37,255.

One complication of the analysis is the fact that, beginning in September 1980 and lasting through 1981, the Guatemalan press ceased reporting incidents of killings and disappearances. The Lucas Garcia government imposed this ban indirectly through its backers in the business community who threatened to withdraw advertising from newspapers reporting human rights violations. The message was reinforced by the deaths or disappearances of seven journalists during the 8 weeks

⁷ The best-known incident of such a delayed execution is that of Dagmar Hagelin, a 17-year-old Swedish citizen who was mistakenly shot during a death squad operation on January 21, 1977. Seriously wounded, she was whisked away to the Navy Mechanics School (ESMA) which also served as a detention center. There she was hospitalized but her injuries left her partially paralyzed. Despite the fact that the incident, which occurred in mid-morning in a suburban neighborhood, was observed by several witnesses, Argentine officials, from President Videla on down, vehemently denied that it had taken place or that Dagmar was in custody. This position was held in face of persistent and energetic inquiries by the Swedish government and many international human rights organizations. Later, she was transferred to Villa Joyosa, another Navy detention center in the city of Mar de Plata where she was relatively well-treated and, although still wheel-chair bound, allowed to roam freely in the villa's wooded grounds. According to one witness, he saw her there in November, 1977—10 months after she had been shot. A Mar del Plata physician has stated that on an evening in June or July 1978 he was called to Villa Joyosa where he was required to sedate a number of prisoners, including a invalid young blonde who he believed was Dagmar. It is currently thought that the sedated prisoners, including Dagmar, were taken to sea and dropped overboard later that night. It is possible that Dagmar herself left a poignant clue to her incarceration in La Joyosa: when her father accompanied by Swedish journalists visited La Joyosa in 1984, they found a tree with the initials "DH" carved on the trunk (Simpson and Bennett 1985, p. 129).

immediately preceding the press blackout. Consequently, there is almost no data for the 16 months when the violence reached its peak.

During the period covered by this survey, contemporary press reports contain 531 incidents in which one or more unidentified victims were killed in the Department of Guatemala. The total number of unidentified killed (UIDK) was 723. The semester values of UIDK victims are plotted against VPAS deaths in Fig 5.10. As indicated by the *coefficient of determination* (r^2) they are significantly correlated at the 0.0001 probability level. The finding supports the hypothesis that many UIDK bodies were buried as *X.X.s*.

Granted that there is a correlation between the UIDK and the VPAS, it is next worth investigating the relationship between the UIDK and the *desaparecidos*. The CIIDH database lists 778 individuals as urban disappeared between 1977 and 1986. The semester values of these variables are plotted in Fig 5.11. The correlation is highly significant ($r^2 = 0.399, p = 0.007$). The outlying value represents semester 84.1 when only 12 unidentified but 110 disappearances were reported.

5.5.6 *Desaparecido Burials in La Verbena*

If an *X.X.* is eventually identified, his or her name is entered as an addendum on the burial certificate. Such cases are rare: of the 3,171 *X.X.s* of our series, only 127 (4.04 percent) were subsequently identified.

Among the identified are 11 individuals, all males, listed as *desaparecidos* (Table 5.9). Five died of gunshot wounds (GSW) to the head, four from cranial

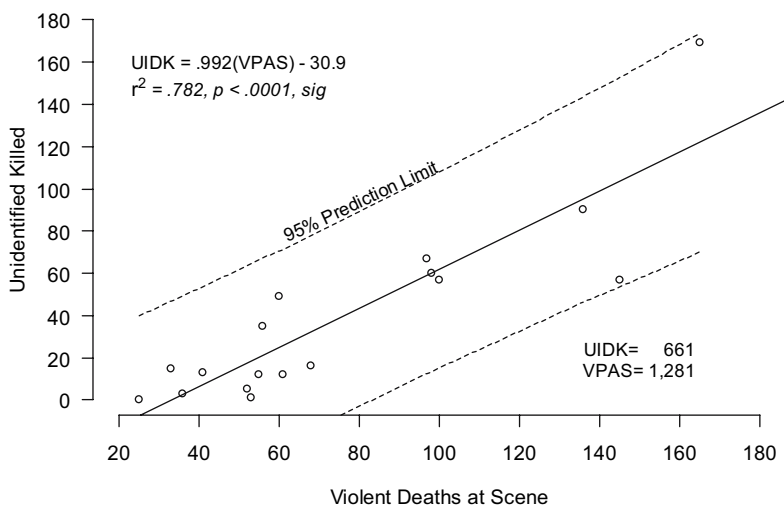


Fig. 5.10 Linear regression of unidentified killed (UIDK) on violent deaths pronounced at scene (VPAS)

Source: CIIDH Database (Ball and Spirer 2000)

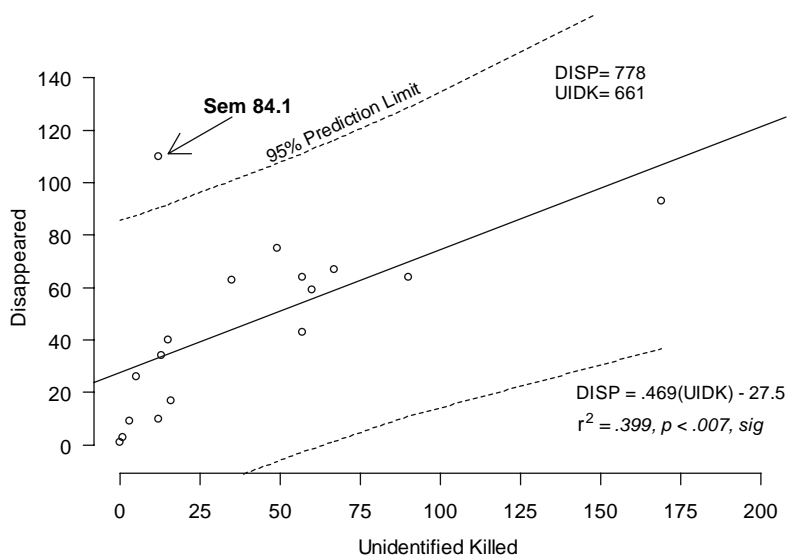


Fig. 5.11 Linear regression of disappeared (DISP) on unidentified killed
Source: CIIDH Database (Ball and Spierer 2000)

blunt force trauma (BFT), one from ligature strangulation, and one from multiple blast injuries received in a bomb explosion. Six cases occurred during the regime of Lucas Garcia and the remaining five during that of Mejía Victores. The bodies of all but one were claimed by their families and exhumed for reburial elsewhere. The burial records offer no clue as to how these bodies were identified or how their families became aware that they had been buried as *X.X.s*.

One of the eleven, Professor Jose Octavio Cajas Sola, was Chairman of the Epidemiology Department at the USAC Medical School and well-known in the relatively small Guatemalan medical community. It is possible that his body was recognized by the examining physician who took it on himself to notify the victim's family of his burial as an *X.X.*

Although their number is small, these victims provide strong *direct* evidence supporting the hypothesis that many of the disappeared were buried as *X.X.s* in *La Verbena*.

5.6 Summary

This study is based on a review of the burial records of *X.X.s* buried in *La Verbena* Cemetery between July 1, 1977 and December 31, 1986. Our findings show the following:

- Between January 1979 and August 1983 (ELEV period) the number of *X.X.* burials rose steeply above baseline values of around 100 per year.

Table 5.9 Disappeared persons buried in La Verbena as *X.X.s* and later identified

Name	Reported missing	Body location	Cause of death	Buried	Exhumed
Oscar Garcia Valazques	Mar-80	Km 28 Carretera A El Salvador	BFI, head	3/2/1980	unknown
Juan Lechuga Santos	Mar-80	6a calle la Av Zona 7 Via Publica	BFI, head	3/2/1980	unknown
Rolando Pinto Olivares	May-80	Residenciales Kananjuyu Zona 16	GSW, head	5/2/1980	10/24/1980
Horacio Mendizabal Garcia	Jan-81	2a. Calle 19-76 Zona 14	GNT, blast	7/26/1981	9/30/1981
Jose Octavio Cajas Sosa	Feb-82	Calz. San Juan 7-12 Zona 7	GSW, head	2/16/1982	4/23/1982
Pedro Sagastume Mendoza	Mar-83	Unknown	GSW, head	12/26/1982	2/3/1983
Lauro Marroquin Guzman	May-84	Calle Mariscal entre 9y10 ave Z. 1	BFI, Head	5/2/1984	Unknown
Baudilio Mendoza Lopez	May-84	Km 65 Ruta al Atlantico	GSW, head	6/17/1984	7/24/1984
Mario Martinez Velasquez	Jun-84	Km 65 Ruta al Atlantico	GSW, head	6/17/1984	6/27/1984
Nery Cardona Orantes	Jun-84	Km 65 Ruta al Atlantico	ASP, ligature	6/17/1984	6/27/1984
Jorge Quintanillas Rodas	Jul-84	5 Ave y 19 calle 19-74 Z. 5, Mixco	BFI, head	7/15/1984	Unknown

Source: GSD.

- The number of *X.X.s* who died of natural causes remained constant and did not contribute to the rise observed during the ELEV period.
- The number of victims who died violently but survived long enough to be pronounced dead at an emergency treatment center also remained constant throughout the study period.
- The increase in *X.X.* deaths during the 4.5 year ELEV was confined to victims who died violently and were pronounced dead at the scene (VPAS).
- This overall increase in VPAS deaths was made up of statistically significant increases in all forms of violence: gunshot wounds, asphyxiations, edged instruments, and generalized trauma.
- From the data, it is estimated that about 889 VPAS deaths in excess of those normally expected occurred during the ELEV period.
- The bodies of the VPAS victims exhibited a statistically significant increase in gunshot wounds to the head during the ELEV period.

- The number of VPAS cases exhibiting multiple traumata from gunshot wounds, asphyxiation, and edged instruments also increased significantly during the ELEV period.
- The number of bodies exhibiting postmortem mutilation rose significantly during the ELEV period.
- During the 3.73-year regime of General Lucas Garcia, VPAS deaths averaged 25.2 per month—a fivefold increase over that of his immediate predecessor. They remained significantly higher than normal during the 1.38-year rule of General Rios Montt.
- As indicated by the steep rise in VPAS deaths, the urban campaign started around November 1978—shortly after a team of Argentine “anti-subversive” operatives arrived in Guatemala to serve as advisors to the Lucas Garcia government.
- The urban campaign ended in the early months of the Mejía Victores. Its termination was marked by a transient rise in VPAS deaths as closing detention centers were “tidied up” by execution of prisoners still on hand.
- VPAS deaths are strongly and significantly correlated ($r^2 = 0.782$, $p < 0.0001$) with killings of persons listed as unidentified in the CIIDH database.
- The correlation between persons listed as disappeared and those listed as unidentified is significant ($r^2 = 0.399$, $p = 0.007$).
- The burial records show that in 11 cases, X.X.s were eventually identified as persons originally reported as *desaparecidos*.

5.7 Conclusions

Between 1979 and 1983 (ELEV period), about 889 VPAS victims in excess of the number normally expected were interred as X.X.s in *La Verbena*. The temporal distribution of these deaths is significantly correlated with the number of unidentified persons killed during the period of this study.

The number of VPAS victims displaying multiple traumata suggestive of torture and postmortem mutilation increased during the ELEV period as did those with gunshot wounds to the head—a frequent method of extra-judicial execution. Among the few cases of VPAS X.X.s subsequently identified were 12 of persons listed as *desaparecidos*.

The increase in VPAS deaths occurred during the intense campaign conducted against the urban guerillas and their sympathizers (real and perceived) during the successive regimes of Generals Lucas Garcia and Rios Montt.

Based on the above conclusions, FAFG has developed a project to exhume a selected series of X.X.burials from *La Verbena*, beginning in 2006. Its principal objective will be to determine whether the skeletal remains are sufficiently well preserved to yield DNA samples that can be used to establish positive identification. If so, a large-scale project to exhume all of the 1977–1986 X.X.s will be conducted.

While the urban campaign was centered in the Department of Guatemala, operations against individual guerillas and their cells were conducted in other departments as well. Many of these victims would have also been buried as unidentified bodies

in local municipal cemeteries in the departments where they were killed. Studies of the cemetery records in these departments should shed light on these cases as well.

It is hoped that, eventually, the mystery surrounding many of the urban *desaparecidos* can be at last resolved. The forensic evidence recovered in this effort will be collected, analyzed, and preserved as evidence in any judicial proceedings that may be brought against the perpetrators. Finally, the bones themselves will be returned to their families to be properly memorialized, thus ending their long years of hiding in plain sight.

References

- American Association for the Advancement of Science (AAAS) (1986) Guatemala: Casos Reportados: 1980–85. AAAS, Washington
- Ball P, Kobrak P, Spierer HF (1999) State violence in Guatemala, 1960–1996: a quantitative reflection. AAAS, Washington
- Ball P, Spierer HF (2000). AAAS/CIIDH database. AAAS Washington
- Black AN, Burns BD, Zuckerman S (1941). Experimental study of the wounding mechanism of high velocity projectiles. *British Medical Journal* 2:872–874
- Comisión para el Esclarecimiento Histórico (CEH) (1999) Guatemala: Memoria del Silencio. CEH, Guatemala City
- Coupland RM, Meddings DR (1999) Mortality associated with use of weapons in armed conflicts, wartime atrocities and civilian mass shootings: literature review. *British Medical Journal* 319:407–410
- Doretta M, Snow C (2003) Forensic anthropology and human rights: the Argentine experience. In: Steadman DW (ed) *Hard evidence: case studies in forensic anthropology*. Prentice Hall, Upper Saddle River, pp 290–310
- Feinstein AR (1988) Scientific standards in epidemiologic studies of the menace of daily life. *Science* 242:1257–1263
- Feldman DL (1985) The United States role in the Malvinas crisis, 1982: misguidance and misperception in Argentina's decision to go to war. *Journal of Interamerican Affairs and world affairs* 27:1–22
- Kobrak, P (1999) *Organizing and Repression in the University of San Carlos, Guatemala, 1944 to 1996*. American Association for the Advancement of Science, Washington
- Lerer LB, Kugel C (1998) Delays in the identification of non-natural mortality. *American Journal of Forensic Medicine & Pathology* 19:347–351
- Piquet Carniero L (2000) *Violent crime in Latin American cities*. University of São Paulo, São Paulo
- Recovery of Historical Memory Project Initiative (REMHI) (1999) Guatemala: never again! Recovery of historical memory project. Orbis Book, Maryknoll
- Sabato E (ed) (1984) *Nunca Más, Informe de la Comisión Nacional sobre la Desaparición de Personas*. Editorial Universitaria de Buenos Aires, Buenos Aires.
- Simpson J, Bennett J (1985) *The disappeared and the mothers of the plaza*. St. Martin's Press, New York
- Snow CC, Bihurriet MJ (1991) An epidemiography of homicide: Ningun Nombre burials in the Province of Buenos Aires from 1970 to 1984. In: Jabine TB, Claude RP (eds) *Statistics in the service of human rights*. University of Pennsylvania, Philadelphia.
- Yamauchi P (1993) Patterns of death: Descriptions of geographic and temporal patterns of rural state terror in Guatemala, 1978–1985. *Physicians for Social Responsibility Quarterly* 3:67–78

Chapter 6

The Demography of Conflict-Related Mortality in Timor-Leste (1974–1999): Reflections on Empirical Quantitative Measurement of Civilian Killings, Disappearances, and Famine-Related Deaths

Romesh Silva and Patrick Ball

Abstract Assessments of the magnitude and patterns of human rights violations have been done by many methods, most frequently using convenience samples of qualitative interviews and multiple systems estimation. In recent work with the Commission on Reception, Truth and Reconciliation in Timor-Leste, estimates of conflict-related mortality were derived using a combination of survey-based estimates and multiple systems estimation based on graveyard records and qualitative testimony data. This chapter presents the data and methods used to estimate conflict-related mortality between 1975 and 1999 in Timor-Leste. A discussion of the comparative advantages and disadvantages of the respective data sources and methods is presented.

6.1 Introduction

Discussions about large-scale human rights violations in protracted conflicts often start with the question of “How many?” In these contexts, social and political debates about the past often turn on whether the extent and pattern of violations were “widespread” (as opposed to being isolated events), or “systematic” (following patterns which are consistent with the hypothesis that such acts were “part of a policy or plan by those responsible”).

However, information about such violations and their victims is often incomplete due to the challenges of gathering such information during the conflict. In post-conflict attempts to document the past, the exact information about the atrocities can be difficult to recall. Those responsible for human rights violations often go to great lengths to cover their tracks, adding another impediment to effective documentation. Access to the affected population may be limited because large parts of that population may be displaced or traumatized as a result of the conflict, which makes it difficult to observe or make contact with the affected population. Even when the affected population is accessible, access to the area and infrastructure can be limited.

The challenge for statisticians and social scientists is to ensure that debates about “How many?” are based on high quality data and scientifically defensible analysis. Informing such debates and discussions with scientific rigor can help to ensure that victims’ stories will be well represented and that those responsible will be held accountable for their actions against established international human rights norms.

In this chapter, we present data and methods which were recently employed in Timor-Leste to measure conflict-related mortality after the cessation of conflict. We show how combining narrative testimony data, “found data,” and survey-based results with available population census information have helped uncover new knowledge about conflict-related mortality in Timor-Leste. Based on our experiences in Timor-Leste, we identify the types of human rights phenomena for which survey methods and survey data are most appropriate and those human rights phenomena and situations for which multiple systems estimation techniques are better suited.

6.2 A Brief Historical Note on Statistical Methods Used in Estimating the Universe of Fatal Violations in Large-Scale Conflicts

Researchers have drawn on statistical and demographic techniques in the past to quantify the pattern and magnitude of fatal violations during armed conflicts. In doing so, they have drawn on indirect demographic methods, sample survey techniques, and Multiple Systems Estimation (MSE).¹

In the aftermath of the Rwandan genocide, despite the uncertainty around the size of the Tutsi population, William Seltzer applied indirect demographic methods to available population census data to estimate that the magnitude of genocidal killings in Rwanda was at least 500,000 (Des Forges 1999, p. 5). Heuveline (1998) applied demographic measurement techniques to pre-conflict census data and post-conflict electoral lists in Cambodia, which were developed by the United Nations, to estimate the population size after the Khmer Rouge regime and the extent of “excess” mortality in the 1970s.² Heuveline estimated a central estimate of 2.52 million excess Cambodian deaths, between 1970 and 1979, suggesting a theoretically plausible range of 1.17 million to 3.42 million excess deaths for the entire decade.

Patrick Ball, Herbert F. Spierer, and Fritz Scheuren (CEH 1999) pioneered the application of MSE to narrative testimony data in Guatemala. They estimated that approximately 132,000 killings (with a standard error of 6,568) were committed in Guatemala between 1978 and 1996. More recently, Brunborg et al. (2003) used MSE techniques in estimating that at least 7,475 people are missing in connection

¹ For a description of the Multiple Systems Estimation (MSE) technique, please see Section 6.4.4.

² Note that the estimate excludes disappearances and killings that occurred in one region where data were inadequate.

with the fall of the Srebrenica enclave on July 11, 1995. These quantitative findings contributed to the finding by the Trial Chamber of the International Criminal Tribunal for the Former Yugoslavia that the Bosnian Muslims of Srebrenica were targeted for genocide.

Both Physicians for Human Rights (PHR) and the US Center for Disease Control (CDC) applied survey-based methods to quantify killings and disappearances in Kosovo (PHR 1999, Spiegel and Salama 2000). The PHR survey of 1,209 Kosovar refugees in 31 camps found that over one-third (35 percent) of survey respondents either witnessed Serb police or soldiers killing someone (14 percent), or saw dead bodies they believed were killed by Serb police or soldiers (21 percent). The CDC study by Spiegel and Salama found that from February 1998 through June 1999, 67 (64 percent) of 105 deaths in the sample population of 1,197 were attributed to war-related trauma, corresponding to 12,000 (95 percent confidence interval (CI) 5500–18,300) deaths in the total population.

Our work in Timor-Leste, on behalf of the Commission for Reception, Truth and Reconciliation (CAVR, by its Portuguese acronym),³ drew heavily on these advances in using demographic and statistical methods to clarify conflict-related mortality between 1974 and 1999. Conflict-related mortality in Timor-Leste during this period consisted of two main components: famine-related deaths and political violence deaths (i.e., killings and disappearances). We estimated that there were at least 102,800 ($\pm 12,000$) conflict-related deaths in Timor-Leste between 1974 and 1999. Of these deaths, approximately 18,600 ($\pm 1,000$) people were killed or disappeared, while the remainder (i.e., 84,200 ($\pm 11,000$) people) died due to hunger and illness in excess of what would be expected due to peacetime mortality (Silva and Ball 2006).

6.3 Challenges in Measuring Conflict-Related Mortality in Timor-Leste

The scale of conflict-related mortality during Indonesia's occupation of Timor-Leste has been the subject of considerable debate: estimates range from a low of 40,000 to more than 200,000.⁴ The most informed observers have repeatedly concluded their analyses by recommending that direct evidence be gathered and analyzed.

³ CAVR was an independent statutory authority set up in 2002 by Regulation 2001/10 of the United Nations Transitional Administration for East Timor (UNTAET). Its purpose was to inquire into human rights violations committed by all sides, between April 1974 and October 1999, and facilitate community reconciliation with justice for those who committed less serious offenses. The Commission was headed by seven National Commissioners, staffed by Timorese and supported by a number of international technical experts. In addition to a national office, CAVR maintained an office in each district of Timor-Leste. The Commission began its work in February 2002 and completed its work in October 2005. Its final report "Chega!" (Portuguese for "no more, that's enough") was handed over to the president of Timor-Leste on October 31, 2005.

⁴ Estimates based on official Portuguese, Indonesian, and Catholic Church data suggest an overall magnitude of approximately 200,000 deaths. See, for example, Kiernan (2003, pp. 585–597) and

For example, historian Robert Cribb suggests five techniques for measuring total deaths:

1. Perpetrators' accounts
2. Counting physical bodies or graves
3. Capturing historical memory through interviews
4. Indirect estimates via census records
5. "Intuitive" estimates projected from informed observers' understanding of local conditions.

Cribb had concerns in 1999–2001 that, at the time that he published a series of papers on the subject, all the available estimates to date, including his own, were limited to the two weakest methods: indirect and intuitive estimates. In his extensive review of estimates of conflict-related mortality in Timor-Leste, Waddingham writes: "We have to concede, however, that it is not yet possible to produce from available evidence, a quantitatively accurate, generally agreed figure on the death toll in Timor-Leste" (Waddingham 2003, p. 2).

In addition, as the 1980 and 1990 Timorese censuses publishes population counts without disaggregating the population data by key demographic variables (such as age and sex), the ability of social scientists to apply standard demographic estimation techniques to the official population data was limited. Non-governmental sources of population information were also of limited value due to the severely restricted access independent monitors and humanitarian groups had to Timor-Leste during the conflict.⁵

Thus, a key challenge we faced in Timor-Leste was the lack of pre-existing population information that could be used to develop scientifically defensible estimates of the number of East Timorese killed during the conflict with Indonesia. In order to overcome this limitation, we developed three new datasets in partnership with the CAVR as a basis for new empirical analysis of conflict-related mortality in Timor-Leste between 1974 and 1999.

6.4 Presentation of Data and Methods

We developed the following three data sources in partnership with CAVR:

1. The Human Rights Violations Database (HRVD): A collection of narrative statements from individuals who were willing to talk to the CAVR,
2. The Graveyard Census Database (GCD): A comprehensive census of public graveyards in each of Timor-Leste's 13 districts.

Gunn (1997) pp. 26. On the lower side, see Cribb (2001), p. 34. Waddingham offers a review of estimates derived from "intuitive" and indirect methods, see Waddingham (2003), p. 14.

⁵ This is in stark contrast to the conflict periods in Guatemala and Perú where NGOs actively collected large-scale human rights data during the conflict which was subsequently used for detailed quantitative analysis after the conflict. See, for example, Ball et al. (1999, 2003).

3. The Retrospective Mortality Survey (RMS): A random-sample household survey used to measure displacement and mortality during the CAVR's mandate period (1974–1999).

These three data sources are described in detail in the following sections.

6.4.1 HRVD

Between February 2003 and March 2004, CAVR collected narrative statements from individuals in all 13 districts of East Timor and from Timorese then living in West Timor. The CAVR established offices in each of the country's 13 districts as part of the CAVR's socialization and reconciliation process. A total of 7,669 narrative statements were collected documenting reported human rights violations. These narrative testimonies provided information on both fatal and non-fatal violations during the CAVR's reference period.⁶ In order to analyze this qualitative information statistically, it was coded into a statistical database using the design standards of the "Who Did What To Whom" data model (Ball 1996). Although those data provide many useful insights, the CAVR statement-taking process that generated them did not employ a probability-based random sample.

6.4.2 GCD

To develop baseline mortality data for East Timor, the Commission undertook a census of public graveyards in East Timor's 13 districts. Through this process, available information about name, date of birth, date of death, and religion was collected. Gravestones that lacked such information were also enumerated and their size was noted.⁷ By collecting this information, the Commission created a *de facto* death registration system for the Timorese population.

A "public graveyard" in this study was defined as a location which is reserved exclusively for burial of deceased persons. This definition includes communal burial sites which are established on public land or land owned by a religious institution. However, it excludes family graves which are built on private property. Enumeration teams documented all gravestones within public graveyards—both marked and unmarked.⁸ A marked grave was defined as having a physical structure which memorialized a person's life, with legible markings in English, Bahasa-Indonesia,

⁶ The HRVD documented 11,126 unique fatal violations and 74,024 unique non-fatal violations.

⁷ Size of an unmarked gravestone can be used as a proxy indicator of whether the deceased was a child or an adult.

⁸ Gravestones of 281,808 were enumerated during the data collection process in 2,582 graveyards. Of these gravestones, 69.4 percent (195,468/281,808) were marked gravestones.

Tetun, or Portuguese.⁹ On all enumerated marked gravestones, the following information was coded if on the gravestone: full-name, date of birth, and date of death. Unmarked gravestones were typically with small simple crosses or other burial markers, without name or date information for the deceased. Enumerators noted information about the religion, type of material and grave size, if it was discernible from the gravestone, for both marked and unmarked gravestones.

6.4.3 RMS

The RMS used standard household survey techniques based on the reported deaths to estimate total number of killings, deaths by hunger and illness, and combatant deaths. However, these totals are estimates of the total number of deaths *that were possible to be remembered by current residents of Timor-Leste*, which is a subset of the total deaths that actually happened. This survey drew a stratified random sample of households, and used a structured questionnaire to collect information about deaths in the family. Deaths preceding the CAVR's mandate were documented in the RMS to aid in the construction of a pre-conflict mortality baseline. However, respondents were only asked about lineal relatives (namely parents, siblings, and children)—as our estimates were developed around the survivorship structure of Timorese families.¹⁰ The sample design of the RMS and weight adjustment calculations used for the RMS-based mortality estimates are described in detail in the final section of this chapter.

6.4.4 Multiple Systems Estimation

Multiple Systems Estimation (MSE) uses several separately collected incomplete lists of the population.¹¹ The lists are matched identifying the elements common across lists in order to estimate the number of elements that are missing from all of the lists. In this project, deaths documented in the HRVD, RMS, and GCD were matched across the three systems using the name, date of death, location of death, and date of birth.¹² Furthermore, sample-survey estimates of these phenomena were also derived using the RMS.

⁹ Due to a lack of resources, we were unable to enumerate Chinese graveyards. The Chinese population in Timor during the Indonesian occupation of Timor has been estimated at 20,000 persons. For analysis about the Chinese population in Timor in the late 1970s, see Kiernan (2004), p. 204.

¹⁰ Estimating mortality from the survey involved estimating the probability of reporting each death based on survivorship.

¹¹ The initial application of multiple-systems estimation to demographic estimation was by Sekar and Deming (1949).

¹² For commentary on the use of these methods in human rights analysis, see Scheuren (2004).

The most basic form of the MSE technique is capture–tag–recapture, which uses only two lists.

A technical explanation of how a count of the unknown members of the population can be estimated using MSE is as follows. Consider the case of two projects P_1 (a list of A individuals) and P_2 (a list of B individuals). There are M individuals who are matched across both lists, in a universe of N total individuals (N is unknown). If all of the people in the universe N have an equal probability of appearing in List 1, then the probability of a specific individual being reported by P_1 is

$$\Pr(\text{Captured in List 1}) = \frac{A}{N}$$

Similarly, if all of the people in universe N have an equal probability of appearing in List 2, then the probability of a specific individual being reported by P_2 is

$$\Pr(\text{Captured in List 2}) = \frac{B}{N}$$

The probability of a specific individual being captured in both lists is

$$\Pr(\text{Captured in List 1 and List 2}) = \frac{M}{N}$$

By definition, the probability of an event composed of two independent events is the product of the independent probabilities. Therefore,

$$\begin{aligned} \Pr(\text{Captured in List 1 and List 2}) &= \Pr(\text{Captured in List 1}) \\ &\times \Pr(\text{Captured in List 2}) \end{aligned}$$

which is $\frac{M}{N} = \frac{A}{N} \frac{B}{N}$. Given this equation, solve for N . Rearranging the terms, and then multiplying by N , multiplying again and finally dividing by M yields $N = \frac{AB}{M}$. Note that with the final equation, the total number of deaths N can be estimated using the totals from A and B and from the matches between them, M .

There are many assumptions implicit in this solution. For example, none of the lists have individuals reported twice and that matching between the lists is accurate. In this project, these two assumptions were controlled during the data processing as described in the matching section.

Other assumptions inherent in the capture–tag–recapture model are more difficult to manage. First, the method assumes that individuals are not entering or leaving the universe during the process of creating the lists, and second that the lists were selected randomly from the population. In human rights documentation projects, the first assumption is usually irrelevant because the documentation occurs retrospectively. The second assumption cannot be satisfied, and it must be replaced by the assumption that the estimation is robust to the selection process.

Another assumption is that the lists are independent; that is, the probability that an individual is in List 2 is independent of the probability that the individual is

captured in List 1. The final assumption is homogeneity: that the individuals that compose the universe all have the same probability of being captured.

If either of these assumptions is violated, the capture–tag–recapture method will not yield an adequate estimate of the total population size. If there are more than two lists with adequate information, the problems of dependency or heterogeneity can often be managed through the specification and selection of appropriate models. However, in the data for the HRVD, RMS, and GCD, there are only two usable systems (RMS-GCD for deaths due to hunger and illness, and HRVD-GCD for killings).¹³ Alone, these estimates would be insufficient, but in combination with the RMS estimates, these estimates provide useful additional information.

6.5 Comparative Analysis of Different Data Sources and Methods

This section presents our four main observations, based on empirical work in Timor-Leste, concerning

1. the applicability of and interaction between survey methods and MSE in measuring large-scale human rights violations, and
2. the strengths and limitations of convenience sample data, “found” data, and survey-based data.

6.5.1 Different Data Sources Have Different Strengths and Limitations

Most human rights projects base their empirical findings principally on databases derived from the large-scale collection of qualitative testimonies. In this, this project in Timor-Leste was no different from the previous projects in Haiti, South Africa, Kosovo, or Perú. In those past projects, analysts were able to draw on substantial additional information that had been collected by governmental and non-governmental human rights projects. Additional sources are important in order to “triangulate,” or understand the patterns and magnitude of human rights events from perspectives other than one particular (qualitative) source. There is something to be gained from every source, and hence we should never discard a data source. But we also need to question every source and test to see what that source is capable of telling us about social reality and, perhaps even more importantly, what that source is incapable of telling us (i.e., what are its implicit biases).

This section discusses the strengths and limitations of the three different data source types used in Timor-Leste: coded-narrative testimonies based on a

¹³ A thorough discussion of the estimators for the dual-system approach and the relevant error calculations is available in Bishop et al. (1975).

convenience sample in the form of the HRVD, “found data” in the form of the GCD, and a household survey in the form of the RMS.

Each data source documents only a small fraction of the total number of deaths in Timor-Leste during 1975–1999. Even in the absence of conflict, not all of the dead are buried in public graveyards: some people are buried in remote locations or in private family graveyards. When mortality conditions are especially severe, relatively fewer people are buried with formal markers. Markers degrade over time, so that by the time the graveyard census was taken in 2003–2004, many graves could not be documented because their information was illegible. Other markers were destroyed entirely in the period between the burial and the time the GCD was collected, leaving only an unmarked gravestone.

The HRVD testimony data was based on a convenience sample, whereby deponents “self-selected” themselves. The style of interviews were open-ended narratives which meant that the breadth of information was wide, but precision of information such as dates, locations, and demographic details of victims was reduced. Furthermore, the transformation of almost 8,000 narrative testimonies (some of which were more than 10 pages long) was labor-intensive and time-consuming. Furthermore, the “self-selection” of deponents, in turn, introduced a number of factors which affected who would be able to give a statement and possible bias, such as the following:

1. People who lived in remote and mountainous areas very far from where the data were being collected (such as district towns) had less chance of being in the sample than those closer to regional towns and district capitals.
2. People who were socially active and/or physically agile were more likely to give statements than those who were sick, elderly, disabled, or traumatized.
3. People who were active in the local community or closely affiliated with local village, sub-district and district officials, and elders were more likely to participate in the socialization process and statement-taking because these local statement-collection efforts were often organized through local village structures and officials.
4. People who died before the Commission was formed did not have an opportunity to tell their stories to the Commission; therefore, events that took place in the past tended to be less frequently reported than more recent events.
5. People with little or no access to the media and mass communications were less likely to know about and therefore approach the Commission.
6. People from constituencies that were hostile to the Commission were less likely to make statements.

The RMS, like the HRVD testimony data, was subject to respondent recall errors, and it was able to enumerate only those deaths which could be remembered by current residents of Timor-Leste. Unlike the HRVD, though, various subject probes were developed through cognitive interviewing¹⁴ of the survey instrument.

¹⁴ Eight cognitive interviews were conducted during the questionnaire development phase. These interviews sought to address challenges faced by respondents in (i) comprehension of the question,

Also, the RMS responses were drawn from a stratified sample of all households in Timor-Leste, and therefore responses were representative of all “rememberable deaths” by current residents. In contrast with the HRVD, the RMS required less time-consuming post-fieldwork data processing. However, it did require intensive work prior to the data collection to prepare sampling frames, design the sampling plan, and design and test the survey instrument.

The RMS reflects the experiences reported in 1,396 households, but the survey omits the experiences of the nearly 190,000 households not sampled. The HRVD reflects the experience of 7,668 respondents, but approximately 940,000 other Timorese did not give testimonies to CAVR. However, even if the HRVD and RMS did reflect the social knowledge of every 2004 resident in Timor-Leste, many deaths would still remain undocumented because all the people who could remember them have died, left the country, or were psychologically or physically unable to recount the stories during the data collection period. In villages where mortality was especially heavy, there may have been no witnesses who survived until 2002–2003. Other families may have left Timor entirely, taking with them their social memory of the deaths. Still other families may have decided to keep secret their past experiences, so it may not be possible to directly document deaths in their family. Social memory is always partial.

Even in the absence of conflict, not all of the dead are buried in public graveyards: some people are buried in remote locations or in private family graveyards. When mortality conditions are especially severe, relatively fewer people are buried with formal markers. Markers degrade over time, so that by the time the graveyard census was taken in 2003–2004, many graves could not be documented because their information is illegible. Other markers were destroyed entirely in the period between the burial and the time the GCD was collected. Nevertheless, the available amount of data from the GCD was still substantially larger than the RMS and HRVD and the documented dates of birth and date from the GCD were not subject to the same level of “recall bias” as the HRVD and RMS.

When comparing these three datasets, we note that their patterns of included and excluded deaths are likely to correlate because conditions that produce non-marked burials and grave loss occurred in the late 1970s, which was 25 years or more before the survey and testimony data were collected. In addition, if people who died were unlikely to be buried in a public graveyard were also unlikely to be remembered by survey respondents, then the two systems would have a positive correlation. Note that this correlation is likely in the immediate post-invasion years

(ii) retrieval from memory of relevant information, (iii) decision processes, and (iv) response processes. Cognitive processes and responses about time and date-related questions indicated that often, when a respondent answered “don’t know,” they may just not have known the exact date according to the Gregorian calendar. However, their responses indicated that sometimes notions of time were easier to recall in relation to major events, or agricultural or seasonal variation. From the cognitive interviewing process, structured date probes were developed which asked the respondent to narrow event-dates into a 6-month window which could be defined by major events such as holidays, or environmental/physical indicators (height of corn or other crops, rainy season or dry season). As a result of the cognitive interviewing, a careful set of probes was created to elicit more detailed descriptions of the places where people were displaced.

when many people were living in very difficult conditions, at first moving constantly and later being held in resettlement camps. Both conditions would tend to lead both to catastrophic deaths of entire groups and also to situations in which relatively few people were buried in public graveyards with permanent markers. The positive correlation between the GCD and RMS in extraordinary years would bias the two-system MSE estimate downward, potentially significantly.

6.5.2 Complementary Methods for Measuring Common and Elusive Phenomena During Conflict

The conflict-related deaths in Timor-Leste included both violent or direct deaths (killings and disappearances) and indirect deaths (resulting from hunger or illness). Deaths resulting from hunger or illness during the early years of the conflict (namely 1975–1979) were common amongst the general population. However, violent deaths were relatively rare amongst the general population, as evidenced by the fact that classical sampling methods used in the RMS only documented 1.5 percent of total killings (as estimated by the MSE). The different data collection methods were complementary in that the RMS was effective in estimating deaths due to hunger or illness but not able to effectively measure the elusive phenomena of violent deaths.¹⁵

As can be seen in Fig. 6.1, the RMS-based killing estimates suffer from an important lacuna in the data: 1991 should have a small peak representing the Santa Cruz massacre (in which qualitative analysts have suggested 270 people were killed and 200 others were missing), but insufficient reports of the event were captured to estimate the killings in that year.¹⁶ The estimate for 1991 is actually slightly lower than the estimate for 1990. This illustrates a challenge with large-scale human rights data collection in conflicts of a similar nature to that in Timor-Leste: killings are relatively rare events across the entire population, so probability samples are unlikely to capture specific events (e.g., there were few reports of the Santa Cruz massacre in the RMS). In qualitative reports such as the HRVD, investigative resources must be specifically devoted to specific events, or there is no guarantee the events will be documented. Also, surveys which use adaptive sampling designs can be used to estimate rare populations and may be more appropriate for measuring elusive phenomena than data collection which is based on classical sampling methods.¹⁷

The HRVD documented nearly one-third of the total killings that were estimated using MSE, whereas as mentioned above, there are only 235 documented killings in the RMS. Consequently, the MSE for killings combines the GCD and the HRVD

¹⁵ For discussion about statistical sampling issues involved in the measurement of elusive phenomena, see Sudman et al. (1988).

¹⁶ For an eyewitness account, see Stahl (1991), Amnesty International (1997), and Human Rights Watch (1991).

¹⁷ For a general introduction to adaptive sampling methods, see Thompson and Seber (1996). Also, for an example of adaptive sampling methods used to measure elusive human rights phenomena, see Jennings and Swiss (2000).

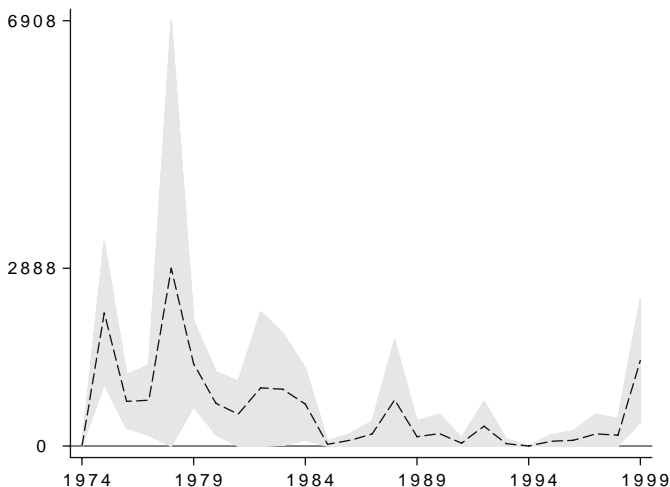


Fig. 6.1 RMS-based estimates of total killings in Timor-Leste over time, 1974–1999
 Source: Retrospective Mortality Survey conducted by CAVR.

data. The results are shown in Fig. 6.2. The MSE estimates 18,600 total killings ($\pm 1,000$). The vertical axis notes the maximum of the error (3,260) and the maximum estimated value (2,634), both of which occurred in 1999. The pattern over time is much clearer in the MSE than in the survey estimate, as can be seen by comparing Fig. 6.1 to Fig. 6.2: the estimated total number of killings rises from nearly zero in

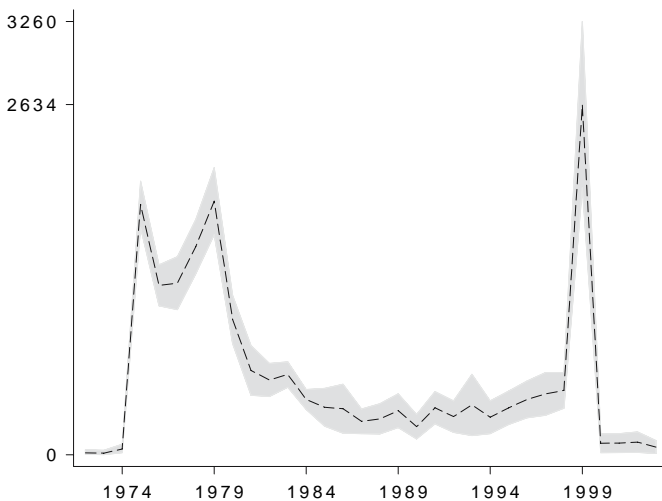


Fig. 6.2 MSE-based estimates of total killings in Timor-Leste over time, 1974–1999, was 18,571 (± 982)

Source: Multiple systems estimation using the Human Rights Violations Database and the Graveyard Census Database created by CAVR.

the pre-invasion period and peaks in 1975 and 1979. Killings decline thereafter and through the 1980s and 1990s. A spike in 1999 marks the high point of estimated killings, significantly greater than any other year.

Although the HRVD received more than 20 reported deaths for the Santa Cruz massacre, this is only a small fraction of the total. Much smaller events were more thoroughly covered, and so the estimates over time do not accurately reflect the importance of that year.¹⁸ Furthermore, the nature of deaths due to hunger and illness seems to have a lower propensity of reporting despite their widespread nature. This is consistent with the hypothesis that violations resulting from acts of omission (such as deaths due to hunger or illness) are less likely to be reported by deponents in open-ended narrative testimonies than violations resulting from acts of commission (such as killings and disappearances).

In contrast to deaths due to hunger and illness, killings are unlikely to be substantially underestimated by a dual systems estimate using the HRVD and GCD. First, killings are less likely to affect entire families than hunger and illness deaths, so there are more likely to be surviving relatives to report these events. Second, the ratio of documented killings to estimated killings (the coverage rate) is 0.637, which is higher than the ratio of documented hunger and illness deaths to estimated hunger and illness deaths (0.513). The higher coverage rate for killings means that MSE itself was able to better correct for the unreported killings than for the unreported deaths due to hunger and illness. On these grounds, we find that the finding for killings be that there were approximately 18,600 killings, with a margin of error $\pm 1,000$.

The MSE estimates for hunger and illness deaths are shown in Fig. 6.3. The total estimated deaths by hunger and illness in 1975–1999 using MSE is approximately 123,500 ($\pm 5,200$).¹⁹ The MSE estimate for hunger and illness deaths is lower than the corresponding survey estimate which is shown in Fig. 6.4, which is consistent with the hypothesis that there is positive correlation between the RMS and the GCD. The median number of deaths due to hunger and illness during “normal” occupation years of 1983–1998 estimated by MSE (3,727) is similar to the value found by the survey (3,632). That is, the survey estimates and the MSE estimates for “normal” occupation years are similar, but the RMS has higher estimates for the extraordinary

¹⁸ Note that the margin of error specifically represents these “holes” in data of this kind. That is, by its nature standard sampling methods only captures information on a small number of the total events. Some large events (such as the Santa Cruz massacre) may be missed. The estimated error of the estimated total number of events is designed to reflect the uncertainty around the estimated total, including the fact that some large events may be missed.

¹⁹ For the MSE, only the named deaths reported in the RMS are included. The RMS sampling weights were not used. The GCD does not include the manner of death, so records from the GCD were allocated to a “cause-of-death” code (namely killing, death due to hunger and illness, combatant deaths, and other deaths) using the following methods: (i) For the 7,117 GCD records which were matched to the HRVD and/or RMS, the manner of death was learned from the HRVD and/or RMS and (ii) for the remaining 82,717 GCD unmatched records, the “cause-of-death” code proportions from the period containing each year was used to allocate the GCD deaths in that year. The margin of error of the allocation was included in the estimated error for the MSE estimates.

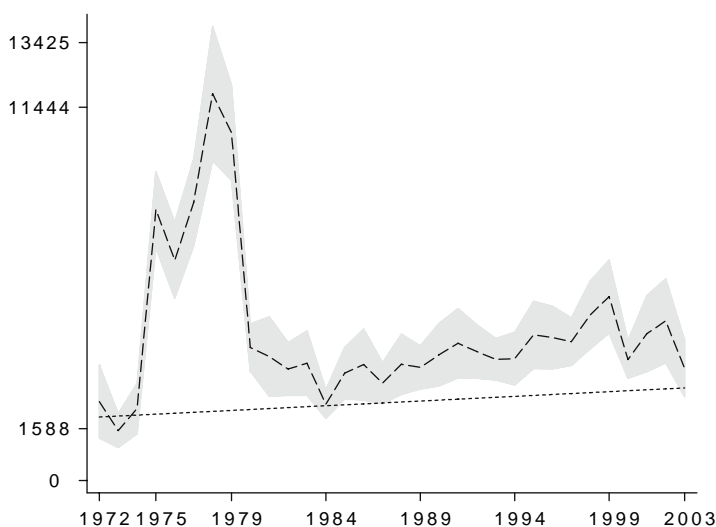


Fig. 6.3 MSE-based estimates of total deaths by hunger/illness in Timor-Leste, 1974–1999. Total estimated deaths by hunger/illness during 1972–2003 was 146,232 ($\pm 5,753$) and during 1975–1999 was 123,529 ($\pm 5,184$). Deaths in excess of estimated baseline during 1975–1999 was 64,037 ($\pm 5,184$)

Source: Multiple systems estimation using the Retrospective Mortality Survey and the Graveyard Census Database created by CAVR.

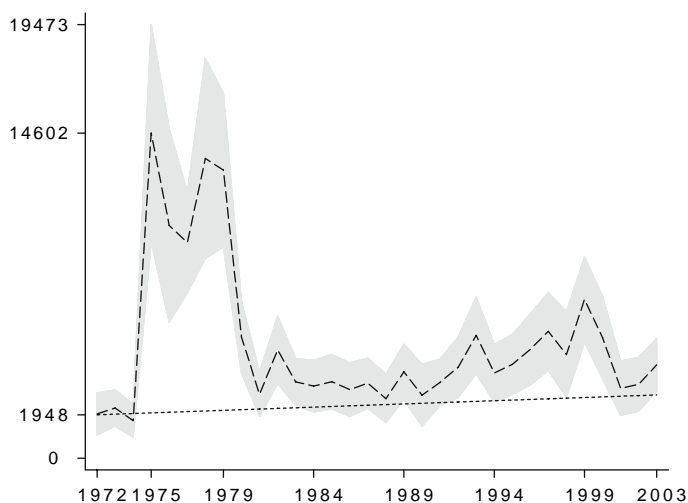


Fig. 6.4 RMS-based estimates of total deaths by hunger/illness in Timor-Leste. Total estimated deaths by hunger/illness during 1972–2003 was 170,127 ($\pm 11,428$) and during 1975–1999 was 143,679 ($\pm 10,925$). Deaths in excess of estimated baseline during 1975–1999 was 84,188 ($\pm 10,925$)

Source: Retrospective Mortality Survey conducted by CAVR.

years. It is likely that during normal years, relatively fewer deaths occur in catastrophic events that eliminate entire families (causing survey underreporting), and relatively fewer people who died are left outside public graveyards. Consequently, during normal years, the MSE and the survey provide similar estimates. During extraordinary years, both methods underestimate total deaths due to hunger and illness, but the MSE underestimates slightly more because a small proportion of all people who died are buried in public graveyards during extraordinary years.

The pattern shown in Fig. 6.3 was similar to the RMS estimate, but the magnitude is lower: note that the maximum estimated value in Fig. 6.3 is 11,444 whereas in the RMS estimate it is 14,602. There was one difference in the patterns over time: the MSE estimated totals for 1975 were lower than the estimates for 1978 and 1979, whereas in the survey the estimates for the 3 years were close together. In the RMS estimate, the error bands for the 3 years were large relative to the differences among the years and, therefore, the hypothesis cannot be rejected that the real totals in those years are equal.

The MSE estimates suggest that 1975 had fewer deaths than implied in the RMS estimates. Other than this difference, the MSE and RMS estimates are similar in magnitude and pattern. It should be emphasized that the two estimates are methodologically very different: magnitude in the RMS was driven by the survey weights, while the preponderance of the data in the MSE came from the number of graves with dates in each year and the matching of the deaths identified in the RMS to the names, dates, and locations of graves.

Combining the results from the two estimates, a highly conservative estimated minimum number of deaths by hunger and illness in excess of the peacetime baseline is between 75,000 and 86,500. These estimates only draw on the 1970 and 2004 censuses and the CAVR's own data. These estimates should be explicitly understood as estimates of the total deaths due to hunger and illness *which were possible to remember in 2004*. This is a significant limitation on the calculations which can be made. Both the RMS and MSE estimates are substantially conservative because many deaths could *not* be remembered in 2004. Some deaths left no surviving family members available to report the death in 2004, and some people who died during extraordinary years were not buried in public graveyards. The years in which the survey is most likely to have been affected by the loss of entire families are also the years in which people are least likely to be buried in public cemeteries. This positive correlation between the GCD and the RMS data creates an underestimation in the MSE.

For killings, the MSE estimates are more appropriate, while for the estimated total number of deaths due to hunger and illness, the RMS estimates are more appropriate. The survey-based estimates from the RMS and the MSE estimates for "normal" occupation years are similar, but the RMS has higher estimates for the extraordinary years. It is likely that during normal years, relatively fewer deaths occur in catastrophic events that eliminate entire families (causing survey underreporting), and relatively fewer people who died are left outside public graveyards. Consequently, during normal years, the MSE and the survey provide similar estimates. During extraordinary years, both methods underestimate total deaths due to hunger and illness, but the MSE underestimates slightly more because

a small proportion of all people who died are buried in public graveyards during extraordinary years.

As discussed above, famine-related deaths (i.e., excess deaths due to hunger or illness) in Timor-Leste during the Indonesian occupation were a widespread, population-based phenomena. Using the RMS, we estimated that at least 84,200 ($\pm 11,000$ people) died due to hunger and illness in excess of what would be expected due to peacetime mortality. According to the US Census Bureau the population of Timor-Leste in 1980 was approximately 557,000 persons (US Census Bureau 2006). In the 1,396 structured interviews which were conducted as part of the RMS, 3,919 famine-related deaths were reported (i.e., 2.8 famine-related deaths were reported per RMS interview). Whereas in the 7,668 open-ended narrative statements which formed the basis of the HRVD, 4,869 famine-related deaths were reported (i.e., 0.63 famine-related deaths were reported per HRVD interview). Given that famine-related deaths were a widespread, population-based phenomena in Timor-Leste, the substantially lower reporting rate of these deaths in the HRVD relative to the RMS suggests that the structured interviewing process used in the RMS was more adept at documenting famine-related deaths than the open-ended interview style used in the HRVD (as convenience sampling alone should not have substantially reduced the HRVD's ability to document population-based phenomena such as famine-related deaths). This finding is consistent with the hypothesis that structured interviewing with customized probes are more appropriate for eliciting information about human rights violations which result from acts of omission, such as deaths due to hunger and illness.

6.5.3 Sample Survey Data can Strengthen Found Data When Estimating Indirect Conflict Deaths Alongside Direct Conflict Deaths

In large-scale conflicts, it is important to distinguish between *direct conflict deaths*, which occur as a direct consequence of fighting, and *indirect conflict deaths*, which arise from the indirect consequences of armed violence, including death from illness, disease, or hunger which would not have happened in the absence of the conflict. A comprehensive assessment of the human rights consequences of mass violence must include the direct deaths from combat or armed violence, and the indirect conflict deaths. In the past, researchers have relied exclusively on epidemiological surveys to measure both direct and indirect conflict deaths. However, the total estimates of direct deaths, using epidemiological surveys, are usually subject to substantially more uncertainty than overall mortality. While the sample size of epidemiological surveys is usually sufficiently large to produce statistically reliable data on crude mortality rates, it is usually not sufficiently large to produce reliable estimates of direct conflict deaths, which themselves are much rarer amongst the general population relative to indirect conflict deaths.²⁰

²⁰ For examples of epidemiological surveys on conflict-related deaths, see Depoortere et al. (2004), Physicians for Human Rights (1999) and Roberts et al. (2004).

By combining survey-based data with found data, researchers may be able to more precisely measure direct and indirect conflict deaths. However, found data projects may only be able to collect partial information about a proportion of the population of interest. Such data may not definitively distinguish between direct and indirect deaths. In Timor-Leste, the graveyard data did not include information about the manner of death. Yet one of the most important research questions in Timor-Leste was “What proportion conflict-related deaths were the result of killings and disappearances, and what proportion were the result of deaths due to illness or hunger?”

Approximately 6.9 percent (7,117/89,894) of the documented gravestones (which had name and date information about deceased) documented by the Graveyard Census Database matched either the HRVD or the RMS (or both) and, through these matches, the manner of death can be learned from the matched record’s manner of death. The remaining GCD records were allocated to the four categories of manner of death (killing, death due to hunger and illness, combatant deaths, and other deaths). Annual proportions of deaths by these four types were calculated using the RMS. These proportions were used to allocate the unmatched GCD records to the distinct manners of death to be used in the MSE calculations for each year: the proportions from the period containing each year was used to allocate the GCD deaths in that year. The resultant proportional allocation of cause-specific mortality is shown in Table 6.1. The margin of error of the allocation was included in the estimated error for the MSE estimates, as is shown in Fig. 6.2 and 6.3.

By allocating unmatched GCD records (whose manner of death was therefore not known) to a manner of death according to the underlying distribution from the survey-data, we were able to make empirical estimates of both direct and indirect conflict-related mortality.

Table 6.1 Estimated proportions of deaths, by period and manner of death

Period	Killing (%)	Hunger/illness (%)	Combatant (%)	Other (%)
1972–1974	0.9	95.9	0.0	3.2
Margin of error	1.8	5.1	0.0	4.9
1975–1982	11.2	83.0	4.4	1.4
Margin of error	4.7	5.1	2.5	0.6
1983–1998	5.5	86.5	0.7	6.2
Margin of error	2.5	3.7	0.6	2.5
1999	16.2	83.0	0.4	0.4
Margin of error	10.2	10.2	0.8	0.8
2000–2003	3.5	86.9	0.8	8.9
Margin of error	3.1	6.5	1.6	4.9
Total	8.3	85.1	2.4	4.3
Margin of error	2.7	3.1	1.2	1.2

Source: Silva and Ball (2006, p. 150).

6.5.4 Sample Surveys Can Help Established Baseline Mortality When Vital Registration Systems Do Not Exist

A key question in Timor-Leste has been “how many people died due to the Indonesian occupation of Timor-Leste? We addressed this question by estimating the excess mortality in Timor-Leste between 1975 and 1999, defined as the additional mortality in the population above that would normally be expected if the pre-invasion mortality rate had continued as it was during the pre-invasion peacetime period.

During the conflict, Timor-Leste did not have a Vital Registration System. Furthermore, the population censuses conducted in 1970 and 1980 lacked the coverage and reliability which would be needed to establish robust estimates of baseline mortality in the 1970s and 1980s.²¹ Hence, a major challenge was establishing the pre-invasion peacetime mortality rate—which is required for analysis of “excess” mortality.

We estimated the total number of deaths due to hunger and illness for Timor-Leste in 1975–1999 using a two-system MSE derived from the GCD and HRVD, as discussed above.²² Some of these deaths are natural in the sense that they would occur in the absence of conflict or famine. An assessment of deaths that could be attributed to the conflict must consider first how many deaths would likely have occurred as a result of hunger and illness in the absence of the conflict.

The immediate pre-invasion years of 1972–1974 provided a peacetime baseline of natural deaths due to hunger and illness. First, in order to create a population baseline, population estimates for 1971–2003 were interpolated between the total reported in the 1970 Portuguese census and the 2004 census total. Then using the RMS, we estimated the number of deaths due to hunger and illness in 1972–1974. Death rates due to hunger and illness for each year were computed by dividing the projected population for each year by the RMS estimated number of deaths for that year.

The estimated death rate was applied to the projected population for each year through 2003 to form the baseline shown. By subtracting the peacetime baseline projection from the annual RMS estimate, we estimated the excess deaths due to hunger and illness. These deaths constitute our estimate from the RMS of the deaths which exceed the total that would be expected if the death rate due to hunger and illness had continued as it was in the pre-invasion peacetime period.²³

²¹ See, for example, Hull (2000), p. 38.

²² The two-systems MSE estimate based on the HRVD and GCD was preferred over a two-systems MSE based on the RMS and GCD for two reasons. First, the number of reported hunger and illness deaths in the RMS in 1974–1999 is small relative to the RMS total estimate of deaths due to hunger and illness. Reported hunger and illness deaths accounted for only 1.6% of the total estimated hunger and illness deaths. Second, two-system estimates can be biased as a result of uncontrolled correlation between the systems.

²³ Both the baseline and the estimated deaths are subject to substantial recall error. In an appendix to our report to the CAVR, we provided a sensitivity analysis that simulated the effect of adjusting the baseline and the total estimated deaths due to hunger and illness to account for the deaths

By using survey-based methods in combination with available population census data, we were able to estimate a mortality rate for the pre-invasion peacetime period and the conflict period. Combining this baseline rate with a two-system MSE estimate of hunger and illness deaths, we were able to establish an estimate of “excess mortality” for the years in which the Indonesian military occupied Timor-Leste (i.e., 1975–1999).

6.6 Conclusions

Sample surveys, multiple systems estimation, and indirect demographic techniques are complementary. Furthermore, these different methodological techniques can leverage wide-ranging data sources (such as narrative testimonies, “found data,” existing population census data, and survey-based data). In certain circumstances, data collected using standard sampling techniques may be more appropriate (such as when estimating conflict-related deaths which are common amongst the general population) whereas in other situations MSE may be more appropriate (such as when the phenomena are less common, even elusive, amongst the general population). Structured questionnaires, which are developed using cognitive interviewing and survey pre-testing techniques, can aide in the establishment of better quality cause-specific mortality data than is sometimes available from open-ended, narrative testimony data and administrative records (e.g., graveyard data). The combined use of diverse data sources and complementary estimation techniques can be adapted to provide researchers a comprehensive means to clarify conflict-related mortality in situations where classical population data systems are damaged, incomplete or non-existent.

6.7 A Note on the RMS Sample Design

The RMS sample was based on a two-stage sample design. The first stage was a sample of all 2,336 aldeias in East Timor, and the second stage was a sample of households within the selected aldeais.²⁴

A total of 1,440 households were originally selected; 1,396 households completed the survey. Households were selected from 138 clusters (made up of aldeias and groups of aldeias). The clusters were selected by a method called “Probability Proportional to Size” (PPS) sampling, and then 10 (or 20) households were selected

which became “unreportable” as all the decedent’s relatives died or otherwise became inaccessible to survey methods.

²⁴ An aldeia is the smallest administrative unit in East Timor. In general, an aldeia is a settlement of group of homes in a small local area. Usually, a suco (village) is made up of three or four aldeias, and groups of sucos make up a subdistrict which is an administrative subset of a district. There are 13 districts, 64 subdistricts, 498 sucos, and 2,336 aldeias in East Timor.

by simple random sampling in each cluster. If each cluster had exactly the same number of sampled households, the sampling probability of each household would be identical, a process known as self-weighting (Levy and Lemeshow 1999).

The population of households was stratified via the following variables: urban/rural, district location, and elevation. Implicit stratification methods were used so that the list of aldeias was sorted by the following ranked variables: urbanicity, district, and altitude, and a systematic random sample picked aldeias across each of the stratification variables.²⁵ A cumulative measure of size variable was created and a sampling interval was calculated as the number of clusters (144) divided by the total measure of size (180,015), which equals 1,250.1. A random number between 1 and 1,250.1 was generated (396.235) and the aldeia with a cumulative measure of size above that number was selected in the sample. 1,250.1 was added repeatedly to the initial randomly generated number and aldeias were selected throughout the list in the same fashion.

Because the sample size within cluster is not constant, the household probability of selection is not constant and an adjustment was made for the cluster size. Non-response was 3.1 percent, and so no non-response adjustment was made. The weights were calculated as follows:

For each cluster, the adjustment for varying cluster size was

$$\text{Cluster adjustment} = \frac{\text{median cluster size}}{\text{cluster size}}$$

The raw 1990 household sampling probability was

$$\begin{aligned} \text{sp}_{1990} &= \frac{\text{total number of sampled HHs}}{\text{total HHs in 1990}} \\ &= \frac{1,396}{168,858} \end{aligned}$$

and so, for each cluster, the PPS weight was

$$\text{pps_wt_raw} = \frac{1}{\text{sp}_{1990}} \times \text{cluster adjustment}$$

There was considerable population change due to migration and growth between 1990 and 2004 when the survey was conducted. Before the weights could be estimated, the total number of households in each aldeia was adjusted from the 1990 census using data from the 2004 census. During the sample design, the clusters were chosen using the household counts for each aldeia reported by the 1990 census. At the time these calculations were done (April 2005), the Census Timor-Leste 2004

²⁵ We used a method known as “Probability Proportional to Size” (PPS) sampling (in this case “size” refers to the number of households and not population, although the two are obviously correlated), a common design in surveys of this kind.

enumeration data were available disaggregated only to the subdistrict level, but not by suco or aldeia.²⁶ Note that the 1990–2004 weight adjustments do not affect the total summed weight, which is fixed at the number of households that existed in 2004. The weight adjustments affect how many households in different places affect the projection.²⁷

Although the 2004 household totals are available from the census at the subdistrict level, the RMS has too few responses at the subdistrict level for the estimates of weights by subdistrict to have adequate data (29 of the 59 sampled subdistricts have fewer than 20 responses). Therefore, the 1990 weights were scaled to the 2004 district totals by the following calculation:

$$\text{District adjustment} = \frac{\text{total HHs in 2004 in this district}}{\text{total 1990 weight in this district}}$$

$$\text{pps_wt_2004} = \text{pps_wt_1990} \times \text{district adjustment}$$

By forcing the weights to match the 2004 census's district household counts, the weights were normalized to sum to the total number of households in 2004 (194,943). The errors given in the results were calculated using Stata's standard survey modules (Stata Corporation, 2003). These modules use the survey design variables (stratum, primary sampling units, and sampling weight) to make weighted estimates of the totals and Taylor-series approximations of the sampling errors. The error estimates assume random sampling with unequal sample weights. This assumption is conservative (i.e., it will tend to underestimate the sampling error) with respect to weights calculated using the PPS methods described above (Brogan 2005).

References

- Amnesty International (1997) Amnesty International Report 1996. Amnesty International, New York (www.amnesty.org/ailib/aireport/ar97/ASA21.htm, accessed April 12, 2006)
- Ball P (1996) Who did what to whom: planning and implementing a large-scale human rights data project. American Association for the Advancement of Science (AAAS), Washington
- Ball P, Asher J, Sulmont D, Manrique D (2003) How many Peruvians died? An estimate of the total number of victims of the internal armed conflict, 1980–2000. AAAS, Washington
- Ball P, Kobrak P, Spierer HF (1999) State violence in Guatemala (1960–1996): a quantitative reflection. AAAS, Washington
- Bishop YM, Fienberg SE, Holland PH (1975) Discrete multivariate analysis: theory and practice. MIT Press, Cambridge

²⁶ See <http://dne.mopf.gov.tl> for the 2004 Population Census data for Timor-Leste.

²⁷ Two subdistricts from 1990 were not listed in the 2004 census results: Fatu Maca in Baucau was absorbed by Baucau subdistrict, and in Oecussi, Pante Macassare B was subsumed in Pante Macassare. For these subdistricts, the number of households in 2004 was estimated by using the proportion of households in the absorbing and absorbed subdistricts in 1990 multiplied by the total in the absorbing subdistrict in 2004.

- Brogan D (2005) Sampling error estimation for survey data. In household sample surveys in developing and transition countries, United Nations Publication ST/ESA/STAT/SER.F/96. Department of Economic and Social Affairs of the United Nations Secretariat, New York
- Brunborg H, Lyngstad T, Urdal H (2003) Accounting for genocide: how many were killed in Srebrenica? *European Journal of Population* 19:229–248
- Commission for Historical Clarification in Guatemala (CEH) (1999) Guatemala: Memory of Silence Tz'ini na'tab'al. Report of the Commission for Historical Clarification Conclusions and Recommendations. UN Office of Project Services, Guatemala
- Cribb R (2001) How many deaths? Problems in the statistics of massacre in Indonesia (1965–1966) and Timor-Leste (1975–1980). In: Wessel I and Wimhoef G (eds) *Violence in Indonesia*. Abera-Verl, Hamburg
- Depoortere E, Checchi F, Broillet F, Gerstl S, Minetti A, Gayraud O, Briet V, Pahl J, Defourmy I, Tatay M (2004) Violence and mortality in western Darfur, Sudan (2003–04): epidemiological evidence from four surveys. *Lancet*, 364:1315–1320
- Des Forges A (1999) *Leave none to tell the story: genocide in Rwanda*. Human Rights Watch, New York
- Gunn G (1997) *Timor-Leste and the United Nations: the case for intervention*. Red Sea Press, Lawrenceville
- Heuveline P (1998) 'Between one and three million': towards the demographic reconstruction of a decade of Cambodian history (1970–79). *Population Studies* 52:49–65
- Hull T (2000) From province to nation: the demographic revolution of a people. In: Fox JJ, Soares DB (eds) *Out of the ashes: the destruction and reconstruction of East Timor*. Crawford House Publishing, Bathurst
- Human Rights Watch (1991) *East Timor: the November 12 massacre and its aftermath*. Human Rights Watch, New York (www.hrw.org/reports/pdfs/i/indonesia/indonesi911.pdf, accessed April 12, 2006)
- Jennings P and Swiss S (2000) Statistical information on violence against women during the Liberian civil war: statistics, development, and human rights. *Proceedings of the International Association of Official Statistics*, Montreux
- Kiernan B (2003) The demography of genocide in Southeast Asia: the death tolls in Cambodia, 1975–79, and Timor-Leste, 1975–80. *Critical Asian Studies* 35:585–597
- Kiernan B (2004) War, genocide, and resistance in East Timor, 1975–99: Comparative reflections on Cambodia. In Selden M and So AY (eds) *War and state terrorism*. Rowman and Littlefield, Oxford
- Levy PS, Lemeshow S (1999) *Sampling of populations*. Wiley, New York, Ch 11
- Physicians for Human Rights (PHR) (1999) *War crimes in Kosovo: a population-based assessment of human rights violations of Kosovar Albanians by Serb Forces*. Physicians for Human Rights, Boston
- Roberts L, Lafta R, Garfield R, Khudhairi J, Burnham G (2004) Mortality before and after the 2003 invasion of Iraq: cluster sample survey. *Lancet* 364:1857–1864
- Scheuren F (2004) History corner. *The American Statistician*, 58:13–15
- Sekar CC, Deming WE (1949) On a method of estimating birth and death rates and the extent of registration. *Journal of the American Statistical Association* 58:101–115
- Silva R, Ball P (2006) The profile of human rights violations in Timor-Leste, 1974–1999. A report by the Benetech human rights data analysis group to the Commission on Reception, Truth and Reconciliation. Benetech, Palo Alto (www.hrdag.org/resources/publications/Benetech-CAVR-statistical-report.pdf, accessed August 22, 2006)
- Spiegel PB, Salama P (2000) War and mortality in Kosovo, 1998–1999: an epidemiological testimony. *Lancet* 355:2206–2211
- Stahl M (1991) Massacre among the graves. *The Independent* (November 17, 1991)
- Stata Corporation (2003) *Stata survey data reference manual, V8*. Stata, College Station
- Sudman S, Sirkin MG, Cowan CD (1988) Sampling rare and elusive populations. *Science* 240: 991–95
- Thompson SK, Seber GAF (1996) *Adaptive sampling*. Wiley, New York

US Bureau of the Census (2006) International database summary for demographic data for East Timor. US Census Bureau, Suitland (www.census.gov/cgi-bin/ipc/idbsum.pl?cty=TT, accessed May 4, 2006)

Waddingham J (2003) Timor-Leste Death Toll, 1975–1999, submission to the Commission for Reception, Truth and Reconciliation (CAVR). On file with the CAVR, Dili (July 22, 2003)

Chapter 7

Afghan Refugee Camp Surveys in Pakistan, 2002

James Bell, David Nolle, Ruth Citrin and Fritz Scheuren

Abstract Both as professionals and as citizens, the events of the last years have brought about many changes to our view of the world and our engagement in it. This survey was one response to those changes. Its main goal was measuring attitudes on a variety of social, economic, and political issues of the Afghan refugees that were returning to their homeland from Pakistan. Particularly important was learning about their perceptions regarding current circumstances as well as future expectations. From a methodological perspective, trying to obtain a good sample of adult males in the refugee camps posed many challenges and most of the discussion will be focused on those challenges.

7.1 Introduction

This chapter provides initial findings from a survey conducted in the spring of 2002 in a sample of Afghan refugee camps in Pakistan.

The survey is an unusual but very important and timely example from the overseas research program that has been mounted historically by the US government, initially by the U.S. Information Agency and now by the Office of Research, Department of State. That Office is the official pollster for the US government abroad.

To orient the reader, we have provided a map (see Figure 7.1) showing the spatial distribution of Afghan refugee camps within Pakistan in 2002. Notice there is a high concentration of camps in the North West Frontier Province (NWFP) bordering eastern Afghanistan.

From the standpoint of its organization, this chapter is divided into ten brief sections, including this introduction. In Section 7.2 the main purposes of the research and the consequent analytic focus are detailed. Next (in Section 7.3) the sample design is described, including how we used the camp lists provided by the United Nations High Commissioner for Refugees (UNHCR) and the limitations imposed by such a frame. The rapid development we had to do of the questionnaire in two languages, Pashto and Dari, is covered in Section 7.4, along with some initial results. Because this population does not have a high literacy rate and is generally unfamiliar with polling procedures, our questions were greatly simplified; consequently, some of the results from this survey probably reflect some acquiescence

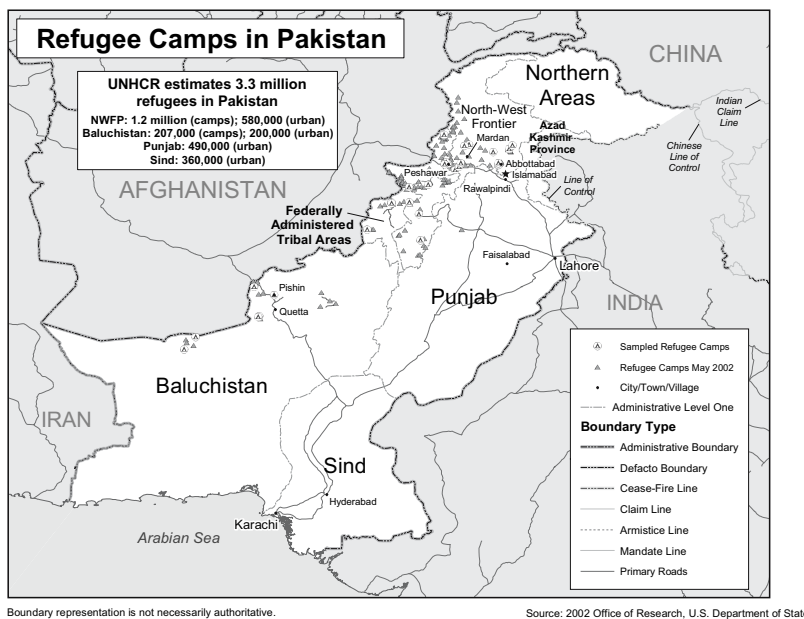


Fig. 7.1 Refugee camps in Pakistan in 2002

Source: Office of the geo-grapher and global issues, U.S. Department of State.

bias (see, for example, Javeline 1999). Section 7.5 describes how we handled various contract survey management issues, some successfully, some less so. Information on the severe design effects that we encountered is presented in Section 7.6. Further analyses are touched on in the next section, Section 7.7, entitled “Next Steps.” Section 7.8 through 7.10 include an afterward and the results from the survey.

7.2 Research Purposes and Consequent Analytic Focus

Because Afghan refugees will eventually take part in deciding Afghanistan’s political and economic future, we wanted, in the survey discussed here, to gauge Afghan refugee attitudes on a variety of social, economic, and political issues. In addition, because we expected that their current circumstances as well as their future expectations might influence their desires to repatriate to Afghanistan, we wanted to get perceptions of their current situations, as well as their expectations regarding their future in Afghanistan.

Those purposes pushed us in the direction of a comprehensive attitudinal survey (see Section 7.4). The design called for coverage of Afghan refugee males as well as females—located in camps as well as in urban areas throughout Pakistan. However, because we anticipated that, for social and cultural reasons, males would

play the dominant role in any decision to repatriate, we focused a major portion of our resources on designing the survey to get a good sample of males in the refugee camps. Only the male refugee sample drawn in the refugee camps is described in what follows.

7.3 Sample Design and UNHCR Frame

To sample males, mainly household heads, in the refugee camps we began with the camp lists provided by the United Nations High Commissioner for Refugees.

This frame was dated and, as we expected, somewhat incomplete. Even so, the UNHCR camp counts at the time of sampling were believed to offer roughly accurate size measures of the total residency in each camp and were relied on for stratification.

Our design was quite standard in most respects but its implementation had many challenges. For frame construction, we had two basic lists: A lengthy list of names and population estimates for 92 older camps, which had been established during various time periods prior to the October 7, 2001 bombing of the Taliban in Afghanistan. There was also a shorter list of eight new camps that had emerged to handle the refugees from the October 7 bombings. In both cases, we wanted to use camps as the primary sampling unit (PSU). Although we recognized that there were problems with the list of older camps, we used this list as the basis for all of our subsequent sampling activities for camps established prior to the October 7 bombings. Consequently, we shall focus on the sampling details related to those two lists for the bulk of the discussion.

At the outset, we allocated 1,200 interviews to the old camps and 200 interviews to the new camps. We anticipated that post-stratification weighting would rectify any disproportionality that we introduced with this initial allocation and our remaining stratification. Because the old camps varied in size from approximately 4,000 to over 75,000, we stratified them into nine relatively homogeneous groups based on rounded estimates of their population sizes as well as their locations. For example, at one end of the continuum in NWFP, our first stratum contained 24 camps averaging around 10,000 refugees per camp, while our second stratum contained 27 camps averaging around 15,000 refugees per camp. At the other end of the continuum, we had three very large camps ranging from 55,000 to 75,000 persons, and we selected each of these with certainty.

Within each homogeneous stratum having multiple camps, we selected the camps with equal probability, and we allocated 40 interviews to each randomly selected camp. We felt that this approach would give us a relatively equal probability of inclusion for all respondents selected within a given stratum. Furthermore, we chose the number of selected camps to correspond approximately to the estimated size of the population in the stratum to avoid any major disproportionalities among strata. In addition, within each stratum, we divided the randomly selected camps into two random halves so that in the end we had 12 camps going into one random half and 12 camps going into the other random half. We also subdivided each of the three

largest camps into random halves where we put 40 interviews into each half, so that we ultimately had two groups of camps each having a total of 15 random halves. We felt that the use of random halves would provide us not only with insurance in the event that the project was disrupted by rapid repatriation or significant shifts in the refugee population but also with reassurance when we analyzed these random halves to see whether they were consistent or divergent in any meaningful way. Furthermore, the random halves for the camps selected with certainty provided us with a basis for our estimation of the within-camp sampling variance in the certainty cases.

7.4 Questionnaire Design and Some Initial Results

Most of the questions in the questionnaire had to be restricted to reasonably simple forms. For example, five-point Likert scales were replaced by simple agree–disagree dichotomies. The amount of work to construct even this type of questionnaire was considerable. The questions had to be in two languages, Dari and Pashto. As a part of our procedures, both the Dari and the Pashto versions of the questionnaire were independently back-translated by Dari and Pashto translators in the US and changes were forwarded to the contractor by fax.

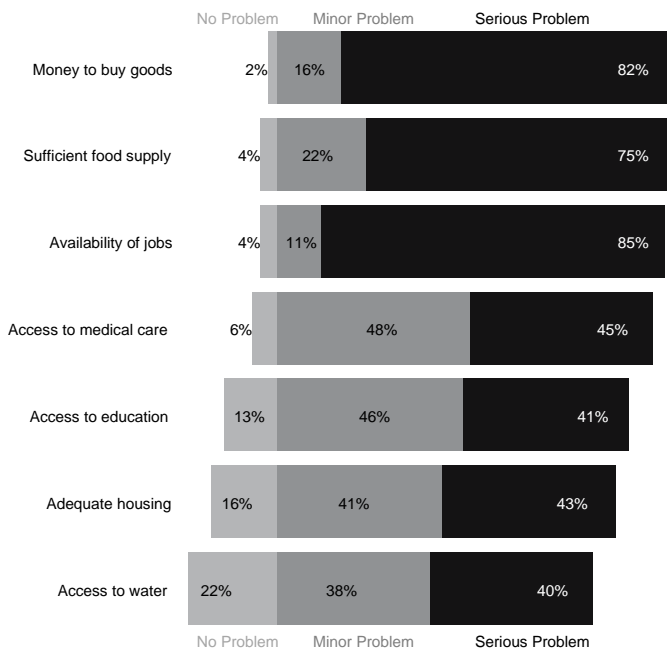


Fig. 7.2 Money, food, and jobs top list of refugee concerns
 Source: 2002Afghan refugee survey by Office of Research, U.S. Department of State

Judging from some earlier related work done by one of our colleagues (Javeline 1999), we knew that there was likely to be some acquiescence bias or “agreement bias” using the agree–disagree approach—that is, there would be an overestimation of agreement with an opinion, say *Opinion A*, when the question is phrased “Do you agree or disagree with *Opinion A*?” Our preliminary analyses suggest that at least some politically oriented questions may have been subject to such bias. However, many of our other questions on social and economic issues and living conditions have different types of question formatting and response alternatives. Thus, it is important to emphasize that the relative ranking of the results shown in the Section 7.10 are probably unaffected by our concern about acquiescence bias. Moreover, many opinions were so nearly universal that we feel comfortable in treating them as reliable enough to act upon.

For example, the bar graph in Fig. 7.2 highlights the respondents’ assessments of their living conditions. Given some of the dire circumstances that are evident in these camps, it is no surprise that access to money, access to food, and the unavailability of jobs top the list of serious problems. It is very significant that even access to a basic need such as water is a serious problem for many respondents.

The full texts for this series of questions as well as other questions used in our initial analyses are reported in the Section 7.10, which also includes the weighted percentages as well as the jackknifed standard errors for those percentages.

7.5 Survey Management

For this survey, the in-country contractor was half a world away from the principal investigators most of the time. In fact, the amount of direct onsite supervision was limited to just one face-to-face set of meetings. Given this situation we had to introduce a number of (remote) control procedures to prevent, if possible, and detect, when necessary, procedural deviations. And we did find some deviations, but they were *not* judged to bias the results to any serious extent, although they did contribute to an increase in the measured sampling error, about which more will be said in Section 7.6. What then did we do? Some examples will have to suffice.

For the pilot survey, the contractor spent one day at a major camp in an effort to pre-test the questionnaire. We also wanted the contractor to check on the usability of the random route approach we were using to get random selections in each camp (see Reed (2002) for related ideas) and to check their skip intervals and turning procedures as they moved along their routes (see the Pashto version of a completed random route contact sheet in Fig. 7.3; note that unique identifying information has been removed). As a result of the pilot, both the questionnaire and the random route contact sheet were revised.

Although the start points and walking directions were not picked by a completely random procedure, our feeling is that any deleterious impacts from this decision may have been minimized by the fact that eight random routes were going to be

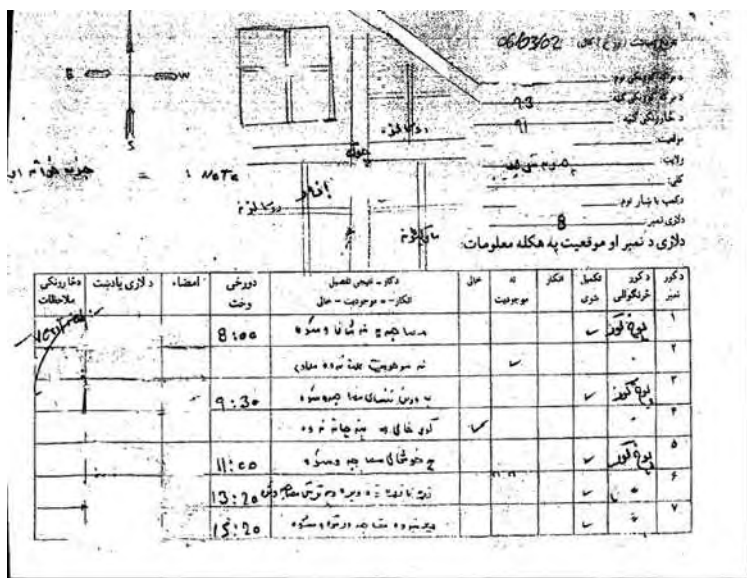


Fig. 7.3 Map used within refugee camp for survey
 Source: 2002 Afghan refugee survey by Office of Research, U.S. Department of State.

assigned in each camp and only five or so interviews were going to be conducted on each route. It was decided that any willing adult male, not necessarily the household head, would be interviewed in each selected dwelling unit—but no more than one per household.

There were a total of ten interviewing teams covering the males in the selected camps. Each interviewing team was composed of a supervisor and two interviewers, and all interviewers were Afghan refugee males. The allocation of the interviewing teams was supposed to follow a pattern so that no one team dominated a single stratum. In eight of the nine strata covering the old camps, this practice held true. We found, by the way, for the eight strata for which we could check that no consistent pattern of difference by team arose. Each of the 20 interviewers ultimately conducted between 40 and 120 interviews in this survey of 1,400 males in 27 old camps and 5 new camps. The average interview was approximately 57 minutes in length. All of those interviews were completed between February 16 and March 22, 2002.

7.6 More Survey Results and Design Effects

The design effects that we encountered in this survey, as already noted, were unusually severe. These effects are presented graphically in the box and whisker plots shown in Fig. 7.4. This graphic presents the DEFT's (that is, the square roots of the

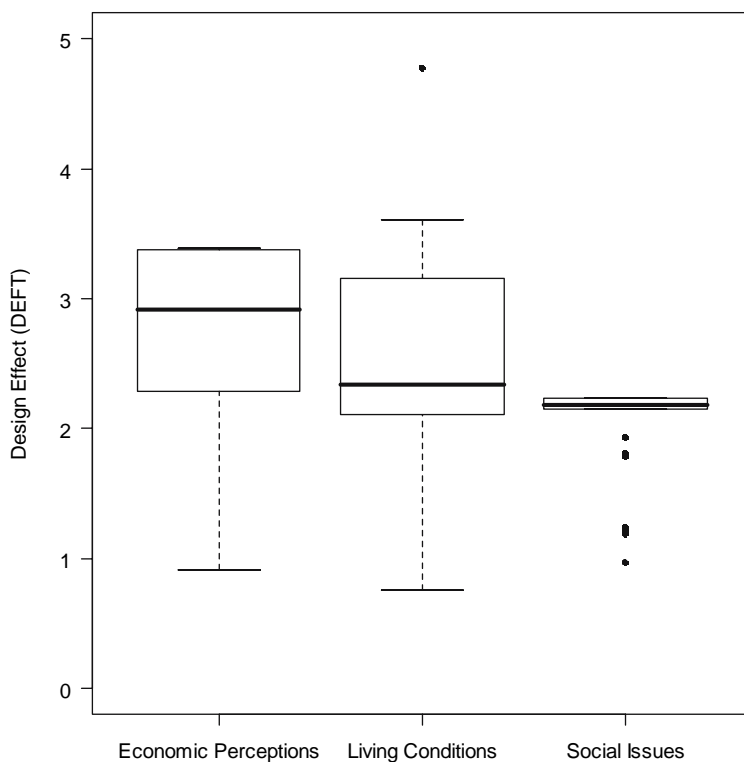


Fig. 7.4 Design effects¹ by question type (Afghan males in refugee camps in Pakistan, 2002)
Source: 2002 Afghan refugee survey by Office of Research, U.S. Department of State.

design effects called DEFF's, see Kish 1965) for all of the response options relating to various types of items initially analyzed in our survey. These DEFT's have been computed from the DEFF's from WesVar 4.0 (see www.westat.com/wesvar/), developed by Westat Corporation.

Each of the DEFT's incorporated into this graphic is a ratio of two kinds of standard errors: The numerator of this ratio is the standard error calculated by a replication method called "Jackknife n" while the denominator is the standard error computed with a simple random sample formula. In all of our preliminary calculations, the finite population correction factor has been ignored.

Through the use of "Jackknife n," our replicates for estimating the error variances in our sample are formed in a way which takes into account our stratification as well as the clustering which arises from the 32 camps used in our male sample. Then our replicate estimates are fed into a generalized formula that ultimately provides the standard error for any given statistic, for example the percentage saying that a

¹A design effect is the ratio of the jackknifed standard error to the simple random sample standard error.

sufficient food supply is a serious problem. The core of that formula is simply the sum of the squared deviations of the replicate estimates of a given statistic from the full sample estimate of that statistic.

Figure 7.4 summarizes the DEFT's for three types of items: economic perceptions, living conditions, and social issues. Three basic things are prominent in Fig. 7.4:

- First, most of the DEFT's are larger than 2.0, which is considerably bigger than DEFT's found in typical national surveys around the world. As a matter of fact, no interquartile range (that is, the outer edges of the boxes) for any of the three types of questions falls below 2.0.
- Second, with the exception of the DEFT's computed on selected social issues, these DEFT's vary considerably within each type of question.
- Third and most importantly, these DEFT's vary substantially by type of question. For example, the median for the DEFT's relating to social issues is slightly above 2.0 while the median for the DEFT's on the three economic perceptions falls close to 3.0.

Those findings have implications for reporting the results from this survey because it is unlikely that a single standard error computed for the total survey will realistically capture the sampling errors generated for these items. In other words, these items will probably warrant special consideration in error reporting.

Meanwhile, that sort of finding (involving large design effects) forces one to search for its source(s). At the outset, we have eliminated case weighting as the key explanation for the large design effects in Fig. 7.4. We discovered that, although the weighted data predictably produce larger DEFT's than the unweighted data, the striking feature is that the increase is very modest: 6.6 percent in the case of the mean DEFT and 3.1 percent in the case of the median DEFT.

Given our analysis of the demographics of some of these camps, we have come to believe that certain common backgrounds, shared experiences, and possibilities for long-term and frequent social interactions have created strong attitudinal clustering, which is likely to be the primary source of these particularly large design effects. In other words, it seems likely to us that most of the design effects for these items appear to be the byproducts of clustering in this sample.

7.7 Next Steps

We are continuing our analysis in several ways. For example, we are reviewing the effects of the presence of others on the interviewee's responses. Section 7.9 of this chapter suggests that these effects are modest for most questions. Detailed analyses of item "nonresponse rates" indicate that for most questions the item "nonresponse rates" were relatively low and comparable between the random halves formed during the initial sampling of the camps.

Meanwhile, further analyses are needed to determine whether the relatively large design effects found for our initial set of items reported herein extend to other types of questions such as ones on social backgrounds and open-ended questions.

7.8 An Afterward

The work described in this chapter was carried out during a period of intense national and international crisis. The scientific reflections made at the time were and are incomplete. We are particularly grateful to the referees for focusing us on broader issues. We cannot thank them enough, even though we have not been able to satisfy them in several respects.

Unapologetically, we have kept to the feel of that time in this chapter—initially written in the summer of 2002. We have made only one significant change: to include all the references mentioned by one of the referees, as good examples of what might well come to several pages of related citations (see Birmingham and Jahoda 1955, Casley and Lury 1981, Fitch et al. 2000, United Nations 2005, Yansaneh and Eltinge 2000). The only other references we added were to our own previous refugee survey work in Albania (Ball et al. 2002) and a general reference to Kish (1965) for those readers who want to know more about design effects—an idea Kish invented. Jackknifing techniques mentioned in this chapter are also covered in Kish (1965).

The previous work of one of the authors (Scheuren) in the refugee camps in Albania (see Ball et al. 2002) did not exhibit the design effects (DEFFs) we found in Pakistan. Perhaps this was because the Albanian camps were temporary. We should have suspected the possibility of greater DEFFs in the long-established camps in Pakistan. We certainly found them. Even so, it was not an option to sample all the camps.

As this chapter tries to make clear, practical issues in the camp selection were severe. In fact, we were not even sure if the UNHCR lists of camps were complete, as events during the winter of 2001 and spring of 2002 were very fluid.

Rough size measures were present but we elected to use these and geographic location to create strata (rather than to conduct PPS camp selections). Thus we ended up, aside from the certainty stratum, with a two-stage design with simple random selection of camps at the first stage and a simple random sample of households at the second. This way resulting weights would be approximately equal within strata and except for the new camps, close to self-weighting across strata.

One of the referees has asked for us to spell out in much more detail some of the techniques we used to mount this very difficult survey. Unfortunately, many of these were only documented at the level we have already provided. For more on the statistical settings that exist in human rights emergencies, see Scheuren (2002).

Nevertheless, even within this highly circumscribed framework, we made additional efforts to use the available material to understand the quality of the survey and various contingencies. For example, data from all 287 random route contact

sheets (see Figure 7.3) relating to the 1,400 males in our Afghan refugee camp surveys were entered into an SPSS data file. Preliminary calculations on this file yielded 1,397 completions (a number which is extremely close to the 1,400 actual male interviews recorded in our primary data file), 116 refusals, 323 not-at-homes, and 57 vacants. The foregoing results implied a response rate of 76.1 percent $1397/(1397+116+323)$, which seems credible in this context. However, these rates did vary dramatically among the 20 interviewers used for our surveys. According to our preliminary calculations on this file, the average (unweighted) response rate by interviewer was 79.1 percent but these response rates ranged from 46.9 percent for one interviewer to 100.0 percent for three interviewers. Some of these differences were probably related to the types of camps that the interviewers encountered.

Moreover, we attempted to disentangle interviewer effects from camp effects on respondents' answers by undertaking a series of statistical analyses that took advantage of our planned assignments of interviewing teams to different sampling strata (see Section 7.5). Tests were derived from loglinear analyses fitting all one-way marginals and all two-way associations involving camps, interviewers, and the selected question's response coded as a dichotomous variable (selected response = 1; other responses = 0). The three-way association between camps, interviewers, and the examined response was omitted from the analysis; consequently, the size of the resulting test statistic was one possible indication of whether the interaction effect of camps and interviewers on the selected response was required to generate an adequate fit for the data. All analyses incorporated structural zeroes to indicate those cases where certain interviewers were not assigned to certain camps; consequently, the complete three-way table in each analysis had only 140 cells that contained data from the respondents. A small adjustment of 0.1 of a case was added to each of these cells not only to distinguish cells having sampling zeroes from cells containing structural zeroes but also to improve the performance of the algorithm in fitting the model to the data. Meanwhile, because the "new" camps had produced many of the large residuals found in our preliminary loglinear analyses, all five of the "new" camps were removed from our subsequent analyses. Furthermore, because the chi-squares were obviously inflated by the design effects, our analytical attention was focused on an Information Index (I-squared) which is analogous to an R-squared suitable for qualitative data.

In an analysis of the respondents' perceptions of their economic circumstances measured by some respondents' saying "bad" in response to Question 2, the result for I-squared implied that a simple additive model focusing on a camp effect and an interviewer effect but ignoring an interaction effect was plausible for these data. In addition, a subsequent set of hierarchical loglinear analyses focusing on a series of I-squares suggested that both the camp and the interviewer had independent effects on the respondents' perceptions of their economic circumstances.

7.9 Survey Paradata

On the following page, data are presented for three of the survey questions.

	Were others present when the interview was conducted?			Did anyone other than the respondent listen to the questions asked?		Were questions answered freely by the respondent?	
Question 2							
Response	Total	Yes	No	Yes	No	Yes	No
Good	3.0%	6.8%	2.7%	7.7%	2.6%	2.7%	7.4%
Okay	38.2%	48.3%	37.5%	47.7%	37.5%	38.1%	39.6%
Bad	58.3%	44.8%	59.3%	44.6%	59.3%	58.6%	53.0%
NR	0.1%	0.0%	0.1%	0.0%	0.1%	0.1%	0.0%
DK	0.4%	0.0%	0.4%	0.0%	0.4%	0.4%	0.0%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
WS	1400	95	1305	94	1306	1330	70
Question 3							
Response	Total	Yes	No	Yes	No	Yes	No
Improve	15.5%	34.4%	14.1%	36.2%	14.0%	15.0%	25.1%
Remain same	17.9%	10.1%	18.4%	9.2%	18.5%	18.2%	11.7%
Worsen	36.3%	25.6%	37.1%	23.7%	37.2%	36.7%	30.0%
NR	1.1%	1.4%	1.1%	1.4%	1.1%	1.0%	3.7%
DK	29.2%	28.4%	29.3%	29.4%	29.2%	29.2%	29.6%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
WS	1400	95	1305	94	1306	1330	70
Question 4							
Response	Total	Yes	No	Yes	No	Yes	No
Better Off	11.2%	17.0%	10.8%	17.4%	10.8%	10.7%	21.7%
About same	8.0%	12.9%	7.7%	12.3%	7.7%	8.0%	7.7%
Worse off	79.6%	67.8%	80.4%	69.0%	80.4%	80.2%	68.8%
NR	0.1%	0.0%	0.1%	0.0%	0.1%	0.1%	0.0%
DK	1.1%	2.3%	1.0%	1.3%	1.1%	1.0%	1.8%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
WS	1400	95	1305	94	1306	1330	70

7.10 Survey Questions and Responses

Note: Dari and Pashto versions of these questions were used in the field. The numbers in parentheses beside the response options are the weighted percentages for the options and their design-based standard errors derived from a jackknife replication method implemented with computer software called WesVar 4.0. An asterisk indicates that no person gave this option. The analytical categorization of each question is noted in brackets after the stem of the question.

Q2. How would you describe your own family's economic circumstances these days—is it good, “so-so”, or bad? (economic perceptions)

Answer	Weighted percentage	Standard error
Good	3.0%	0.6%
Okay	38.2%	2.8%
Bad	58.3%	3.0%
No response	0.1%	0.1%
Don't know	0.4%	0.2%

Q3. A year or two from now, do you expect your family's economic circumstances will improve, remain about the same, or worsen? (economic perceptions):

Answer	Weighted percentage	Standard error
Improve	15.5%	3.1%
Remain about the same	17.9%	2.6%
Worsen	36.3%	3.7%
No response	1.1%	0.7%
Don't know	29.2%	4.1%

Q4. All things considered, do you feel your family is better off, about the same, or worse off than you were when you lived in Afghanistan? (economic perceptions):

Answer	Weighted percentage	Standard error
Better off	11.2%	2.8%
About the same	8.0%	1.3%
Worse off	79.6%	3.6%
No response	0.1%	0.1%
Don't know	1.1%	0.3%

Q5. I am now going to read you a list of issues related to the situation of Afghan refugees. For each one, please tell me if it is a serious problem, a minor problem, or no problem at all for your family? First take “access to water,” is this a serious problem, a slight problem, or no problem at all for your family. Now take. . . (repeat for each item on list)

A. Access to water (living conditions):

Answer	Weighted percentage	Standard error
Serious problem	40.4%	6.3%
Minor problem	37.8%	4.1%
No problem	21.8%	3.8%
No response	*	*
Don't know	*	*

B. Availability of jobs (living conditions):

Answer	Weighted percentage	Standard error
Serious problem	84.6%	1.7%
Minor problem	11.0%	1.7%
No problem	4.4%	0.9%
No response	*	*
Don't know	*	*

C. Sufficient food supply (living conditions):

Answer	Weighted percentage	Standard error
Serious problem	74.8%	2.5%
Minor problem	21.6%	2.2%
No problem	3.6%	0.7%
No response	*	*
Don't know	0.0%	0.0%

D. Access to medical care (living conditions):

Answer	Weighted percentage	Standard error
Serious problem	45.1%	3.1%
Minor problem	47.8%	3.3%
No problem	6.2%	1.5%
No response	*	*
Don't know	0.9%	0.5%

E. Sufficient income to buy goods (living conditions):

Answer	Weighted percentage	Standard error
Serious problem	81.6%	2.2%
Minor problem	15.9%	2.3%
No problem	2.0%	0.4%
No response	0.1%	0.1%
Don't know	0.4%	0.2%

F. Access to education (living conditions):

Answer	Weighted percentage	Standard error
Serious problem	41.4%	2.7%
Minor problem	45.5%	3.1%
No problem	13.1%	2.6%
No response	*	*
Don't know	0.1%	0.1%

G. Adequate housing (living conditions):

Answer	Weighted percentage	Standard error
Serious problem	42.5%	4.6%
Minor problem	40.9%	4.3%
No problem	16.4%	3.6%
No response	*	*
Don't know	0.3%	0.1%

Now, I'd like to ask you a few questions on some issues important to Afghans.

Q6. Thinking about education for Afghan children, do you believe that Afghan girls should or should not have the same educational opportunities as boys?² (social issues):

Answer	Weighted percentage	Standard error
Girls should have the same educational opportunities as boys	87.3%	2.0%
Girls should not have the same educational opportunities as boys	11.5%	1.5%
No response	0.2%	0.1%
Don't know	1.0%	0.6%

Q7. What do you think about Afghan women working outside the home to help support their families? Do you think women should or should not be allowed to work outside the home? (social issues):

Answer	Weighted percentage	Standard error
Women should be allowed to work outside the home	81.5%	2.2%
Women should not be allowed to work outside the home	16.6%	1.9%
No response	0.4%	0.2%
Don't know	1.6%	0.6%

Q8. Sensitive question whose wording is not reported but whose percentages and standard errors are used in the analysis (social issues):

Answer	Weighted percentage	Standard error
Serious problem	79.9%	2.3%
Minor problem	19.3%	2.3%
No response	0.1%	0.1%
Don't know	0.7%	0.3%

Acknowledgment The views expressed herein are those of the authors, and these views do not necessarily reflect the views of the National Opinion Research Center (NORC), U.S. Department of State or the US government. The authors wish to acknowledge the very helpful feedback from the attendees at our presentation before a joint meeting of the Washington Statistical Society and the local chapter of the American Association of Public Opinion Research, excellent comments made then and later from Thomas Jabine, and the comments of two anonymous referees.

²This question, unlike some others, seemed to have had responses sensitive to the presence of others during the interviewing. When others were present, answers consistent with traditional Muslim female roles appeared to be more likely than when others were not present.

References

- Ball P, Betts W, Scheuren F, Dudokovic J, Asher J (2002) Reexamining the correlation between Kosovar killings and refugee flows. American Association for the Advancement of Science, Washington
- Birmingham WB, Jahoda G (1955) A pre-election survey in a semi-literate society. *Public Opinion Quarterly* 19:140–152
- Casley DJ, Lury DA (1981) *Data collection in developing countries*. Oxford University Press, London
- Fitch DJ, Molina Ro, Barrientos T (2000) Weighting of survey data: is the view that such should not be attempted in developing countries reasonable? Proceedings of the section on survey research methods of the American Statistical Association, pp 764–769
- Javeline D (1999) Response effects in polite cultures: a test of acquiescence in Kazakhstan. *Public Opinion Quarterly* 63:1–28
- Kish L (1965) *Survey sampling*. Wiley, New York
- Reed H (2002) Demographic assessment techniques in complex humanitarian emergencies. National Academy Press, Washington
- Scheuren F (2002). Survey sampling technology in human rights research. Proceedings of the section on government statistics and section on social statistics of the American Statistical Association. Published previously in *Amstat News*, 306, 77–78. Also available by looking up the social statistics section at the www.amstat.org
- Scheuren F (Forth coming) Conducting human rights surveys in conflict Zones. To be presented at the 2008 February Meetings of the American Association for the Advancement of Science in Boston.
- United Nations (2005) Draft version of household surveys in developing and transition countries: design, implementation and analysis (<http://unstats.un.org/unsd/HHsurveys/index.htm>, accessed February 24, 2005)
- Yansaneh IS, Eltinge JL (2000) Design effect and cost issues for surveys in developing countries. Proceedings of the section on survey research methods of the American Statistical Association, pp 770–775

Chapter 8

Metagora: An Experiment in the Measurement of Democratic Governance

Jan Robert Suesser and R. Suarez de Miguel

8.1 Introduction

Today more and more institutional actors are turning toward applied statistics and quantitative indicators as a means for developing an evidence-based assessment of human rights implementation, democratic processes, and improvement of governance. A notable set of human rights institutions, international organizations, development agencies, national policy makers, and civil society are looking for reliable figures and robust analysis that will empower their work in the fields of monitoring, reporting, advocacy, or policy design. Even if their growing interest for quantitative methods is neither universally shared nor clearly defined with regard to the scope and effective application of those methods, their interest attests to an increased awareness of the need for proper and robust tools that would enhance the reliability, the comprehensiveness, and the efficiency of existing reporting and monitoring mechanisms. Such an increased awareness is leading to new and promising forms of collaboration between statisticians, human rights practitioners, and institutional actors—a collaboration that marks a significant difference from the reluctance and lack of communication that prevailed during last decades.

Indeed, the feasibility and relevance of “measuring” human rights, democracy, and governance have long been controversial both in the human rights community and also in the international statistical community. The potential of statistical analysis for enhancing rigor and reliability of reporting on human rights was first evidenced by pioneering studies and work undertaken in the 1980s, in particular by David Banks, Richard Claude, Thomas Jabine, and Herbert Spirer, as well as by a series of successful projects conducted in the 1990s in different countries by the statisticians of the American Association for the Advancement of Science (AAAS) and Human Rights Program.¹ Nevertheless, it was only in 2000 that the issue was broadly debated on the occasion of the Montreux Conference on Statistics, Development and Human Rights, attended by policy analysts, human rights practitioners, professional statisticians, and governmental officials from 123

¹ See Ball (1996), Banks (1986, 1989, 1999), Claude and Jabine (1986), Jabine, Thomas B. (1985, 1989), Jabine and Claude (1992), Jabine et al. (1995), Spirer (1990), and Spirer and Spirer (1993).

countries and 35 international organizations.² As a consequence of that conference, a multi-disciplinary North/South network emerged, expanded, and gave impetus to a series of international workshops and consultations on the expected tangible follow-up to the Montreux “operational conclusions.”³ This process led to the design of Metagora, an international project focusing on methods, tools, and frameworks for enhancing evidence-based assessment of democracy, human rights, and governance. The project was hosted by PARIS21, within the Organization for Economic Co-operation and Development (OECD) and, thanks to the support of a community of institutional donors,⁴ it started operations in February 2004. Even though Metagora was limited in scope—involving six projects during this pilot phase—it successfully reached its ambitious goal of designing, organizing, and implementing all planned field operations within a year. It was then able to produce, review, and deliver its first results in a Metagora Forum held in Paris in May 2005. Final reports on the pilot phase of Metagora have been presented by the end of 2006. A second phase for the project has found initial financing for the period 2007–2010.

This chapter focuses on the context and policy incidence of the Metagora experience of “measuring human rights.” It intends to (i) describe the specific Metagora approach and method of work (in Sections 8.2 and 8.3); (ii) comment on some significant lessons of the project and on a few perspectives that have been opened by it (in Sections 8.4 and 8.5); (iii) advocate for a strong integration of the human rights and democracy dimensions into governance indicators and discuss the factors that lead to successful integration of those dimensions (in Sections 8.6–8.9); (iv) elaborate on how and why international aggregate indicators and national statistics-based indicators are complementary (in Section 8.10); and (v) briefly point out some strategic steps that must be taken in the short and medium term (Section 8.11).

8.2 A Decentralized Laboratory for the Development of Methods and Tools

Metagora both undertook several pilot activities and also contributed to three ongoing programs in order to develop and test measuring methods and monitoring tools in different regions of the world. Those included

² The Montreux Conference, held on September 2–8, 2000, was organized jointly by the Swiss Development and Co-operation Agency and the Swiss Federal Statistical Office, and was convened under the aegis of the International Association of Official Statisticians (IAOS), a section of the International Statistical Institute.

³ The operational conclusions, as well as all invited and contributed papers of the Montreux Conference, are available at www.portal-stat.admin.ch/iaos2000/index2.htm

⁴ The pilot phase of Metagora (February 2004–December 2006) was implemented thanks to the generous financial support from the European Union, France (via the Ministry of Foreign Affairs), Sweden (via the Swedish International Development Cooperation Agency), Switzerland (via the Swiss Agency for Development and Cooperation and the Ministry of Foreign Affairs), and Canada (via the Canadian International Development Agency).

- pilot sample surveys on key national human rights issues (such as irregularities, abuses, and ill-treatment in Mexico City, indigenous people's rights in the Philippines, and human rights dimensions of land reform in South-Africa);
- use of existing official household surveys to obtain data on democratic participation and governance issues (in eight African countries and in five South-American countries);
- matching quantitative official data and qualitative information to build target indicators and dynamic databases (on the right to education in Palestine);
- the establishment of a controlled vocabulary and harmonized procedures allowing local NGOs to collect and analyze sensitive data (on human rights violations in Sri Lanka);
- the realization of a worldwide survey on local, national and regional initiatives aiming at measuring democracy, human rights, and governance;
- the production of training materials and their dissemination through an on-line tool.

All of those activities were developing simultaneously and in an interactive and coordinated fashion. The whole project has therefore been implemented as a “decentralized laboratory”—a model of work that Metagora intends to expand in the future via the integration of new initiatives that are both aimed at enhancing evidence-based assessment of human rights and democratic governance and also would contribute to and benefit from the international Metagora community.

The materialization of Metagora's working program relies on a multi-disciplinary and inclusive group of organizations and individuals that regularly shared their experiences, mutually reviewed work in progress, and provided each other with mutual support. Several Partner Organizations⁵ have led pilot activities in select countries, coordinated the interventions of other national organizations and local experts, and conducted regular consultations with stakeholders. The following are for example:

- Over the course of the pilot activities, the Partner Organizations in Mexico, the Philippines, and South Africa met together twice for multi-day workshops in which they reviewed and critiqued each other's work, including questionnaires, sample design, and field operations.
- In the Philippines, for the survey on indigenous land rights, the National Commission on Human Rights worked closely with the National Commission on Indigenous Peoples to organize multiple meetings with leaders of indigenous communities to both develop buy-in and also use local expertise to improve the quality and relevance of the pilot survey.

⁵ Metagora's Partner Organizations are Développement: Institutions et Analyses de Long terme (DIAL, Paris); Fundar, Centro de Análisis e investigación (Fundar, Mexico City); the Human Science Research Council of South Africa (HSRC, Pretoria); the Palestinian Central Bureau of Statistics (PCBS, Albireh/Ramallah); the Philippines National Commission on Human Rights (CHR, Manila); the Secretaría General de la Comunidad Andina (Lima); and the Science and Human Rights Program of the American Association for the Advancement of Science (SHR-AAAS, Washington).

- In Palestine, a Steering Committee was set up to involve a large range of civil society organizations and academic/research centers in the improving of the quality of the data collected by local NGOs and academic researchers. The Palestinian Central Bureau of Statistics held two week-long training courses free of cost: a general review of human rights statistics developed from both quantitative and qualitative sources, and a specific workshop on developing a controlled vocabulary for the mining of qualitative statements for quantitative data.
- All of the Partner Organizations participated in a working group on the Metagora training materials, developing a website on which a general manual, an encyclopedia of terms, and example documents and tools have been mounted.

A cross-fertilization process has therefore materialized both at the local level—among experts and practitioners from academic, governmental, and non-governmental organizations involved in the pilot activities—and at the global level, among all partners and experts contributing to the implementation and assessment of the different actions and phases of the project.

While Metagora relies on the support of several researchers and research institutions (such as AAAS, the French DIAL, or the South African Human Science Research Council), it is not being conducted in the form of or with the criteria of academic research. It certainly mobilizes existing expertise and, by vocation, it aims at applying and disseminating scientific knowledge, robust statistical methods, and proper tools. However, it aims neither at producing *per se* new scientific knowledge or new statistical methods nor at being a forum for scientific exchanges (as this already exists, namely within the framework of the International Statistical Institute or within the American Statistical Association). The real vocation of Metagora is, on the one hand, to enhance evidence-based monitoring of policies based on the work conducted by statisticians, analysts, and stakeholders to work together with proper methods and, on the other hand, to ensure that expertise of those who have strong quantitative skills is transferred to those who have less.

8.3 A Bottom-Up Approach to the Design and Construction of Indicators

Metagora aims to improve the conditions for and the means by which information on democracy and human rights is produced, and to integrate that information into the monitoring of governance. This project uses a *bottom-up* approach consisting of

- together with stakeholders in each pilot country, identifying key human rights, democracy, and/or governance issues for which evidence-based assessment is highly relevant;
- applying statistical methods and tools to the particular national context;
- assessing those methods for their capacity to provide policy-relevant results;

- providing stakeholders with a shared knowledge on the policy issues at stake; and
- drawing the “universal” lessons from selected “local” experiences and formulating recommendations and guidelines for further application of the tested methods elsewhere.

For example, the pilot project on land reform in South Africa was coordinated by the Human Sciences Research Council, but involved input from key stakeholders including government departments, the Human Rights Commission, NGOs (such as the National Land Committee and its affiliates), organized farmers (such as the white farmers), and academics, as well as the National Statistics Council and Statistics South Africa. The project’s primary goal was the development of a survey methodology process by which South Africa’s land reform process could be assessed via the collection of information on the experiences of the people that land reform intends to assist. Samples of respondents were taken in black urban areas, in traditional rural areas, and from farms (both black farm dwellers and also white farm owners). Different populations were interviewed using different questionnaires.

Analysis of the resulting data suggests that land reform issues, even if high on the political agenda, are not highest priorities for any respondents, and that unemployment, poverty, and HIV/AIDS are more pressing concerns. However, few respondents demonstrated that they understood the components of the land reform policy, thus education on land reform should be a policy focus in the future. Due to the success of this survey, which shows measurement related to land reform policy is possible, HSRC currently plans to extend the survey to other areas of South Africa as well as to other African countries.

To summarize, Metagora looks at measuring major national issues in order to produce relevant indicators that represent the effective implementation of human rights realization and democracy as inherent dimensions of governance. At the same time, Metagora realizes the aim of local actors to inform national public debates and national policymaking processes. Even so, the project must favor strong synergy at the international level. In particular, Metagora wants to contribute to work that aims to both provide global information products at the international level and also information reflecting the specific national context. We will come back later to such an ambitious goal.

8.4 First Significant Lessons

At the present stage of its implementation, in the autumn of 2006, Metagora has already reached crucial targets and delivered a number of significant lessons that should be further deepened and discussed within the international statistical community, as well as within the human rights community. Among its main accomplishments, five merits need to be mentioned in the context of this book.

8.4.1 Data on Human Rights, Democracy, and Governance Can Be Collected and Indicators Produced That Are Central to Policy Makers' Decisions

Well-established statistical survey methods of sampling and field operations can be applied to measure human rights, democracy, and governance. As it is the case in all surveys that intend to capture sensitive or complex data, proper design and accurate tests of questionnaires are basic preconditions for obtaining reliable relevant data—this is of particular relevance to ensure that a survey appropriately addresses the issues at stake within the specific political, social, and cultural environments of a particular nation. Although non-response has been, and still is, a serious problem, further study and articulation of assumptions and models can improve statistical estimates. Metagora surveys were designed and implemented as pilot experiences, within a short time frame and with few resources—as a consequence there were a number of sampling and technical limitations on the work. In spite of this, the sample surveys conducted in Mexico, the Philippines, and South Africa are providing significant information on the nature, dimension, and magnitude of the issues at stake.

The surveys' preliminary findings—such as the high incidence of irregularities, abuses, and ill-treatment in Mexico City, the centrality of protection of ancestral land for the effective implementation of indigenous people's rights in the Philippines, or the expectations expressed by the South-African black population with regard to land reform—are particularly relevant for addressing key governance issues, for confirming or correcting stakeholders' perceptions and policy assumptions, and for informing appropriate design of policies and programs aimed at redressing rights and enhancing accountability.

8.4.2 Quantitative and Qualitative Data Can and Should Interrelate to Properly Inform Assessment of Democracy, Human Rights, and Governance

First outcomes of the project show not only that assessment of democracy, human rights, and governance can build on the solid rock of proper quantitative reporting, but also that design and use of measurement methods and tools in those areas must be informed by accurate qualitative research and documentation on the situation and perceptions of target populations, as well as on the assumptions and expectations of all kinds of stakeholders. Qualitative information is essential not only to ensure proper design of survey questionnaires, but also to focus statistical analysis on relevant issues and to provide appropriate contextual frameworks for an effective policy-oriented interpretation of quantitative data.

In the Metagora pilot activities, the design of survey questionnaires and databases was based on systematic records of qualitative information gathered through in-depth narrative interviews with victims of rights infringements, focus group discussions with target populations, substantive reports of local experts, as well as

large consultations and topical discussions with many relevant stakeholders and academic specialists. For example, in Mexico a series of interviews with victims of irregularities, abuses, and ill-treatment by police and police officers helped the survey designers understand the types of irregularities, abuses, and ill-treatment that occurred, the mentality of the police officers involved in the irregularities, abuses, and ill-treatment, and the language used to describe particular interactions between the police and the public. In the Philippines pilot activity, a series of focus group discussions were implemented simultaneously with a random sample survey; the qualitative information from the focus groups helped in the interpretation of the quantitative information gathered via the survey. Additionally in the Philippines, meetings with leaders in the indigenous communities were held to discuss the content and implementation of the survey on indigenous land rights.

As a consequence of those activities, it is now possible to relate the surveys' results jointly with qualitative and contextual information in order that policy-oriented reports of the activities have a solid, comprehensive base. Moreover, the pilot experience of building statistically friendly series of data from records of narrative reports on human rights violations, as well as the attempt to match quantitative and qualitative series of human rights data in a single database, are providing substantive lessons with obvious universal scope.

8.4.3 Official Statistical Agencies Can Be Efficiently Involved in the Measurement of Human Rights, Democracy, and Governance

Contrary to an inextricable prejudice, still too broadly shared within the international official statistical community, Metagora's pilot activities prove that national statistical offices (NSOs) can conduct sample surveys on key human rights, governance, and democracy issues, or provide qualified technical assistance and field logistics to other governmental agencies or human rights institutions responsible for monitoring the implementation of rights, democracy, or governance, or even develop measurement programs jointly with research centers and/or civil society actors.

As factual evidence of the feasibility of measuring democracy and governance issues with official statistical tools, survey questionnaires on those issues have been attached as a supplement to regular household surveys conducted by 11 NSOs in Francophone Africa between 2001 and 2004 and in the Andean region between 2002 and 2005. The supplemental questionnaire focused on subjective and objective measures of poverty, governance, and democracy. For example, in the area of governance and democracy, a general question on how well the government administration is run was followed by specific questions on corruption within the government, support of economic policies, and the respondent's personal definition of democracy. Those questions generally yielded low non-response rates.⁶

⁶ See Herrera et al. (2005).

The Francophone Africa/Andean region experience develops and promotes a very promising tool, as using well-established official surveys presents obvious advantages: the size, the quality of the sampling frames, the effectiveness of proper data collection, the analytical potential of the data collected, as well as the moderate additional costs incurred by adding a specific questionnaire to existing surveys. Analysis of the resulting data is enriched with information on respondents collected from both the regular and also the supplemental questionnaires, making it possible to focus on governance issues in relation to major social and development problems such as poverty and exclusion. In Peru, the more advanced country within the framework of this pilot experience, the survey on “*participación ciudadana*” is today at its fourth round, thereby establishing a...time’. Is this Ok? establishing a routine measurement tool that now allows the undertaking of analysis over time.

In terms of a universal lesson, the activities in Francophone Africa and the Andean region—as well as in Palestine, where the Metagora activities were conducted by the Palestinian Central Bureau of Statistics, and the Philippines, where the National Commission on Human Rights coordinated activities with the National Statistics Office, the National Commission on Indigenous Peoples, and the National Statistical Coordination Board—show not only that NSOs can contribute to substantial enhancements of democracy, human rights, and governance assessment, but also that in doing so NSOs reinforce their specific role, their technical independence, and their capacity to interact as a qualified partner with other public institutions and with organized civil society.

The decision to embark an official statistical agency on a measuring exercise focusing on sensitive human rights or democratic governance issues must take into consideration two criteria: the legitimacy and the efficiency. The legitimacy of official statistics fundamentally resides on the level of public trust in the agency that is responsible for generating and disseminating the information. Such a trust is normally based on the recognized technical skills and professional standards of the agency, as well as on its technical independence that is reflected by independent (not politically influenced or biased) information. In this sense, legitimacy results not only from the history of the agency, the public perception of its work, but also from its effective capacity to work with recognized and respected technical independence when it comes to measuring sensitive issues. In a given national context, the official statistical agency can be widely considered as legitimate, whereas in another context the agency may be perceived as “too close” to the political authorities or potentially vulnerable to pressure. One might assume that this latter scenario is particularly likely in countries with poor human rights and democratic governance records. In such circumstances, it seems unlikely for the statistical agency to positively contribute to the assessment and monitoring in these fields. However, as mentioned earlier, the Metagora experience in French-speaking Africa shows that other factors can palliate weak legitimacy (for instance, a regional or supranational dimension of planned household surveys to which the module on democratic governance can be appended). Moreover, it is worth noting that when an official statistical agency is successfully involved in evidence-based assessment of democratic governance, this might contribute to positively change its image in the public at large—thus increasing its public legitimacy.

The criteria of capacity relates to the suitability of an institution to undertake proper collection of data, statistical analysis, and dissemination of results in a given context. Tools and know-how for designing samples, translating the issues at stake (i.e., the interests and preoccupations of various stakeholders) into clear and pertinent survey questions, as well as having proper field staff are among the assets which can make it possible for an institution to take on responsibility for data collection. In each country, beyond the official statistical agency, different institutions (private and public) may, at a given time, present the skills required. The question to be raised is, therefore, “what makes (or does not make) the official statistical agency the most suitable institution to carry out measurement of human rights or democratic governance issues?”

When tackling the issue of the potential advantages of involving official statistical agencies in evidence-based assessment of human rights, democracy or governance, an additional element merits further consideration. The involvement of statistical agencies in this kind of assessments will have as a consequence, for the information produced, to benefit from the credential or label of “official statistics.” This may not be of any importance for some stakeholders, but others would consider that, thanks to this label, the information produced will be more apt than other kinds of data to hold the attention of government itself. The value added of the label “official statistics” merits therefore to be duly appreciated—nevertheless it should not be over-valued: some experiences have shown that although official statistical agencies collect relevant data and produce significant information on democratic governance issues, the local (national) dissemination of the latter should not be taken for granted. Indeed, institutions may be slow to release survey results owing to the fact that parts of government regard them as sensitive. In this respect, unfortunately the 2-year Metagora pilot phase does not yet provide enough experience to fully inform on the issues related to the dissemination of results.

At the end of the day, it is up to the top management of each official statistical agency to evaluate if and how the involvement in evidence-based assessment can be feasible, suitable, and fruitful in the specific national context. The Metagora community strongly recommends that the decision to be (or not to be) involved in these assessments is taken within an agency’s global strategy for reinforcing its legitimacy and capacities.

8.4.4 The Complementary Role of Statistical Methods to the Work of Human Rights Institutions

Current human rights monitoring mechanisms are mainly based on reporting of individual cases (or series of cases) to human rights institutions (HRIs), as well as on judicial decisions. That form of monitoring is certainly invaluable for purposes of advocacy on individual cases, but it does not provide relevant information on the real magnitude and trends of major human rights issues considered as collective, social, and political phenomena. As HRIs are primarily concerned with the promotion and

protection of human rights on the basis of international norms and standards, they are more aware than any other institution of the lack of reliable information on the magnitude of large human rights–relevant phenomena. However, HRI agents are often unfamiliar with quantitative analysis and therefore tend to consider statistics as an overly reductive approach unable to capture the multiple complex dimensions inherent to their human rights work.

The Metagora pilot activity in the Philippines, conducted by the Commission on Human Rights, has proven that working mechanisms involving different institutions with the appropriate substantive and technical skills can successfully implement measurement of complex problems such as the implementation of indigenous people’s rights. Although that activity was conducted by the Commission, it was implemented in close collaboration with the National Statistical Co-ordination Board and the National Commission on Indigenous Peoples. The survey was designed with the assistance of experts from NSCB and implemented with the support of the National Statistical Office and the Statistical Research and Training Center. Heads of all these institutions gather together in a “Metagora Advisory Council” which was the national policymaking body of the project. Experts of the mentioned institutions, together with representatives of academic circles and the civil society (including indigenous people’s organizations, NGOs, individual human rights practitioners, etc.) form the “Implementing Group of Experts” for the project.

The Philippines pilot experience shows that a strong leadership of HRIs, political commitment of relevant institutions in charge of protection of target populations, proper technical assistance and field support from official statistical agencies, and input from all concerned stakeholders constitute essential assets for the success of that kind of initiative. That experience also showed that a long process of dialogue and mutual learning is required to overcome prejudices as well as unfamiliarity of the various actors with the approaches and skills of the others. The results of that activity prove that, on the basis of a shared analysis, statistical information can effectively complement regular reporting of HRIs and can powerfully inform relevant recommendations of HRIs to executive and legislative powers, in particular on the realization of economic, social, and cultural rights.

In a different context, the pilot survey on irregularities, abuses, and ill-treatment in Mexico City also provides evidence of the complementary role of statistical methods to the work of HRIs not only in identifying specific forms and mechanisms of massive human rights violations, but also in constructively addressing governmental policies and programs aimed at enhancing governance structures and measures to eliminate and prevent such violations. The study—even if limited to a restricted geographical urban area—measured for the first time the magnitude of a phenomenon of major political and social concern. Based on the results of the study, the research team was able to recommend the use of control mechanisms such as radios or surveillance cameras in patrol cars to inhibit the high level of abuses in patrol cars. The Metagora project team also recommended better education of individuals regarding their rights and obligations, transparency measures (such as the rigorous wearing of identification by officials), and public information campaigns.

8.4.5 Statistical Methods Can Substantially Enhance the Research and Advocacy of Civil Society Organizations in the Fields of Human Rights and Democracy

Following a road opened by successful projects carried out in particular by US statisticians in several countries during the 1990s, Metagora embarked on supporting activities aimed at enhancing capacities of civil society organizations (CSOs) to develop and use quantitative approaches and statistical tools in their human rights research and advocacy work. In Sri Lanka, thanks to the program support of the Asia Foundation and to a generous contribution of the Danish Development Agency (DANIDA), technical expertise was provided to the Human Rights Accountability Coalition (HRAC), a network in which different CSOs join efforts to systematically collect, collate, analyze, and share data on human rights violations. Under the HRAC umbrella, the Consortium of Humanitarian Agencies (CHA) focuses on experiences of internally displaced people, the Home for Human Rights (HHR) studies legal case files, the Forum for Human Dignity (FHD) focuses on reports of deaths, and the Institute of Human Rights (IHR) studies records from prison medical clinics. The objective of the HRAC is to develop a massive, objective, and undeniable statistical record allowing CSOs to inject scientifically rigorous evidence into the search for truth about the patterns, magnitude, and responsibility for past violations connected to Sri Lanka's ethnic and political conflicts, as well as to maintain a record of current abuses to assist in monitoring compliance with commitments made in the peace process.

Through HRAC, technical principles were established via the use of harmonized forms to register events and an agreed-upon vocabulary on human rights violations. Categories of violations—including forced displacement, killings, and physical assaults—were developed and defined in detail in order to assist in quantifying the abuses experienced. Those measures ensured standardized coding and processing of comprehensive physical records (archives of CSOs, each containing tens of thousands of individual cases of human rights violations). On that basis, the Sri Lanka activity developed and expanded further to ensure rigorous data processing and quality control and to build an impressive electronic record of human rights violations. That demonstrates that well-established methods for ensuring harmonized data collection and coding contribute to capacity building of CSOs and open promising perspectives for independent and proper evidence-based analysis of massive human rights violations.

In a different context, Palestinian academic research centers and CSOs are providing series of data to the Central Bureau of Statistics (PCBS) for the development of a database on human rights and democracy that contains data from different sources. In its current pilot phase, the database focuses only on the implementation of the right to education.⁷ The database will progressively be extended to cover other

⁷ Examples of the types of data in the database include counts of teaching staff, students, and classrooms by type of school (governmental, private, etc.), region, supervising authority, and gender; and university and college enrollment by gender.

human rights areas and issues. Today, in its prototype form, the pilot database is providing a common basis for gathering, coding, and analyzing sets of information from different sources that otherwise may have remained fragmentary, anecdotal, and irrelevant for comprehensive analysis. As in the case of Sri Lanka, the pilot activity in Palestine confirms that appropriate training on data collection techniques and processing of data, as well as adequately calibrated transmission of know-how, substantially empowers the monitoring and advocacy capacities of CSOs.

8.5 The Inventory of Initiatives

Initiatives of various natures aimed at measuring democracy, human rights, and governance are emerging around the globe. What is the scope, relevance, methods, and outcomes of those initiatives? To answer that question, Metagora launched an ongoing world-wide survey on those initiatives, with a specifically developed questionnaire that aims at identifying the location of the initiative, initiating institutions, topics covered, and sources of funding.⁸ The survey has a particular emphasis on initiatives being carried out at the local, national, and regional levels. The preliminary results reveal the existence of numerous ongoing initiatives around the world that were, until now, often unknown by the human rights and statistical communities. Also, the survey is providing information on international initiatives not yet included in the Essex/Eurostat database or in the United Nations Development Programme/Eurostat repository of “sources of governance indicators.”

The preliminary results of the survey show that many of the initiatives are relevant for informing policies and are largely convergent with the Metagora objectives. As a matter of fact, the survey confirmed the need for more rigorous use of quantitative methods both in the collection and analysis of existing data and also in producing new information.

Henceforth, Metagora will interact with many of the institutions and organizations identified in the Inventory of Initiatives. As such, those institutions and organizations will be asked to join a network that consolidates know-how on methodologies for the collection and use of governance information.

8.6 The Current Context for the Production and Use of Indicators of Democracy and Human Rights

As we mentioned before, in the year 2000 the International Association for Official Statistics (IAOS) and the Direction de la Coopération et du Développement (DCD) organized the Montreux Conference on Statistics, Human Rights and Development. For the first time, statistical institutions, international organizations, and human rights advocates discussed their concerns within a dedicated forum.

⁸ The questionnaire form for the Inventory of Initiatives is included at the end of this chapter; results of the survey are available at the Metagora website (www.metagora.org).

The participants reported the hopes and concerns of the communities they represented. The hopes were for the timely availability of indicators. The concerns were centered on the potential misuse or even manipulation of these same indicators.

Six years later, those hopes and concerns endure. The need for indicators for quantifying democracy and human rights is prominent. Examples are many. The European Union (EU) specifies respect for human rights as a condition for benefiting from its external assistance programmes; therefore, the EU is in dire need of tools for quantifying human rights. The mission of the United Nations Human Settlements Programme (UN-Habitat) is articulated on the notion of (human) rights. Transparency International aims at integrating the many dimensions of corruption into one cohesive concept. More and more, democracy and human rights are seen as fundamental for good governance.

The clear need for a measurement tool does not prevent resistance and controversy. As we saw in the case of the Human Development Index championed by the United Nations Development Programme (UNDP), many questions arise: What is the value of an indicator? Can it be used for practical purposes? Should we generalize all comparisons and classifications between countries on the basis of that indicator? For national and international policy makers and for human rights advocates, the stakes could not be higher.

Among statisticians involved in public policy, a debate rages: Is the danger of using statistics to measure what is often considered a political phenomenon too high? Would the use of statistics as a basis for policy development incur the risk of “politicizing” statistics and inviting partisanship?

Governance’s complexities now cover many areas. Existing pseudo-indicators have hit boundaries. Work to improve their accuracy and extend their scope has received considerable support through a program instated by Daniel Kaufman and his team under the auspices of the World Bank Institute. But their work also suggests limited availability of bona fide indicators and the insufficient local adaptive qualities of those indicators that are available.

In spite of the controversy, the international sphere has never before been so favorable for factoring the notions of “democracy” and “human rights” into governance indicators. It is in that context that methodological efforts can be undertaken, as is the case for Metagora, to produce reliable quantifiable information disseminated as widely as possible by diversified agents operating within the debate on democracy and policymaking.

8.7 Governance, Democracy, and Human Rights: The Trump Cards of the Government

Today, democracy and human rights are essential components of governance, which is, itself, an essential dimension of policymaking related to development.⁹ When populations take ownership of policymaking in its application, the social and economical benefits are considerable. The stronger their insistence in the areas of

⁹ See, for example, Commission of the European Communities (2005).

democracy and human rights, the deeper is the engagement of the individuals in their society. Participation in debates, decisions, and politico-social changes has an important impact on an individual's perception of the public policy. Such active participation opens new perspectives and generates new expectations. Conversely, policymaking itself can be conducted in such a way as to create changes in the expectations of a population.

We can illustrate this process by a concrete situation: Public opinion in the Madagascan capital, as measured by DIAL, is no longer that "corruption is a calamity pervading all governmental institutions and that is normal because power and corruption go hand-in-hand"; that stance has transformed into the belief that "corruption is unacceptable, and to purge governmental institutions from corruption, we the citizens and our political representatives must change our behavior."

8.8 Including the Human Rights and Democracy Dimension in the Indicators of Governance

Databases providing human rights measurements developed by international institutions in recent history remain centered on civil and political rights. The underlying approach is strictly normative. Extending that approach to more concrete policies would require also considering economic, social, and cultural rights.

The Metagora effort is an instance of a more comprehensive approach. It leads to an account of human rights specific to each situation. Investigations conducted in the context of Metagora in Mexico, South Africa, and in the Philippines, as well as investigations on democracy and governance coordinated through DIAL and conducted by national statistical institutes in Francophone Africa and in the South American Andean community, are illustrations of what can be done in that direction. Those experiences led to a better understanding of the strengths and weaknesses of national public policy when they

- produced relevant indicators reflecting the interests of the national stakeholders regarding the link between democracy, human rights, and governance;
- successfully collected quantitative information through the various interested players;
- analyzed a mixture of the available quantitative and qualitative information;
- allowed for the use of the indicators to enlighten decisions; and
- constructed a methodology meeting the expectations of the national political stakeholders, civil society, and the international community.

8.9 The Utility of Catalysts of Confidence and Expertise

Governance indicators depend on the collection of quantitative and qualitative information. When it regards democracy and human rights, such information is especially difficult to capture in virtue of its high sensitivity at both the national and also the

local levels. Only information proven to be relevant for the various stakeholders will contribute to governance in a positive manner. The confidence generated by its production and dissemination is indispensable.

To overcome technical and political difficulties, indicator development must have a repository of resources such as Metagora available; such a repository is a decisive trump card for success. It is important that those resources are not affiliated with a partisan institution and that they include multi-disciplinary expertise.

Such a repository of resources does not serve a substitute for national stakeholders. Rather, it facilitates the process of construction and dissemination of information (both from a technical and from a political standpoint). It also supports the process of dialogue, technical collaboration, and sharing analysis at the national level. The objective is to encourage partnerships between the national and local stakeholder institutions. The ultimate goal is to make endogenous the production of sensitive information through the conduct of pilot experiments.

8.10 Global and National Indicators of Governance

During the last 10 years, all kinds of indicators of governance have been proposed. The international community has been seeking valid indicators that are at the same time simple and versatile. Specifically, several proposals were for global indicators that can be used for evaluating, and even ranking, the performance of national governments. For example, Transparency International produces an annual TI Corruption Perceptions Index and has also created a Bribe Payers' Index that ranks corruption by source country and industry sector,¹⁰ and the Human Development Index of the United Nations Development Programme, also annually produced, strives to compare nations as to the health, education, and employment of their citizens.¹¹

At the same time, research on indicators of governance supported by the World Bank Institute led to a new paradigm in which the potential uses of governance indicators are expanded. The May 2005 report, "Governance Matters IV," states that

while . . . aggregate governance indicators have been useful in providing a general snapshot of the countries of the world for various broad components of governance . . . they remain a rather blunt instrument for specific policy advice at the country level. As we have argued in past, these aggregate indicators need to be complemented with in-depth in-country governance diagnosis, based on microsurveys of households, firms and public officials within the country. The lessons being drawn from these combined aggregate and micro-data sets do point to the importance of moving concretely to the next stage of governance reforms, in Africa and elsewhere.

¹⁰ See www.transparency.org/tools/measurement (accessed September 18, 2006).

¹¹ See hdr.undp.org/reports/global/2005/pdf/HDR05_backmatter.pdf (accessed September 18, 2006), p. 333.

Also, during the Metagora Forum (May 24–25, 2005), Daniel Kaufman of the World Bank Institute specified

... On in-depth in-country diagnostic, it is extremely important, in tailoring to the country needs, and to the issues at stake, to ensure that the issues which are sometimes a bit lost in the aggregate indicators, specialized issues like the indigenous community, how they are treated, the rural areas, what's happening in the decentralized local communities—those can be covered by appropriate tailoring on the specific in-depth efforts and it is a unique opportunity to do so. ... By contrast with the issue of corruption, for instance, thanks to the efforts of Transparency International and others, we have tried hard also, there has been an enormous progress in measurement of corruption over the past ten years. But that has not been matched in every other area of governance. I am sure you have discussed a lot one area that stayed behind, very much compared, say, with corruption, or with government effectiveness, is the area of human rights.

To summarize, the types of indicators suitable for comparison at the international level are not refined enough to inform national policy decisions even when, via comparisons with other countries, they are capable of indicating policy problems. Indicators developed from initiatives such as the Metagora project, which include democracy and human rights concepts in their measurement of governance, are better able to inform national policy, but may not be appropriate for cross-country comparisons. As such, international aggregate indicators and national statistically based indicators are complementary—using both, one can formulate nation-specific policy recommendations within an international context.

8.11 Future Work

The relevance of statistical methods for the production of reliable information in the arena of democracy and human rights to support governance indicators is no longer in doubt. Investigations on democracy, such as those conducted by DIAL,¹² show the full spectrum of successful institutional possibilities. The investigations conducted under the auspices of Metagora in Mexico, South Africa, and the Philippines provide irrefutable evidence as to the value of quantitative data for informing public debate.

There is much left to be done. Some possibilities for the future are

- validating the construction of quantitative time series indicators, allowing for their regular use in the context of national qualitative information regarding governance, democracy, and human rights;
- promoting through collaboration, for example with human-rights institutions, the institutional appropriation of the indicators at the national level;
- testing the collection of information within a regional approach for indicators tied to political governance; and
- obtaining consistency between global and national indicators.

¹² DIAL: Développement et insertion internationale.

The objective then becomes being able to provide partial answers to the following three questions.

1. How should we proceed to use reliable quantitative information on democracy and human rights as a basis for the debates and political decisions aimed at improving governance?
2. How should we proceed to replicate the same methodological model as we created via Metagora for measuring democracy and human rights in several countries?
3. How should we proceed to ensure indicators are used in a relevant and acceptable way by stakeholders who may not share the same views or have the same interests?

References

- Ball P (1996) Who did what to whom? Planning and implementing a large scale human rights data project. American Association for the Advancement of Science (AAAS), Washington
- Banks D (1986) The analysis of human rights data over time. *Human Rights Quarterly* 8: 654–680
- Banks D (1989) Patterns of oppression: an exploratory analysis of human-rights data. *Journal of the American Statistical Association* 84:674–681
- Banks D (1999) The collision of government, ethics, and statistics: case studies and comments. *Proceedings of the Social Statistics Section of the American Statistical Association*, pp 167–172
- Claude RP, Jabine TB (1986) Symposium: statistical issues in the field of human rights. *Human Rights Quarterly* 8:551–699
- Commission of the European Communities (2005) Proposal for a joint declaration by the Council, the European Parliament and the Commission on the European Union Development Policy (COM (2005) 311final) (http://ec.europa.eu/comm/development/body/development_policy_statement/index_en.htm, accessed September 3, 2006)
- Herrera J, Razafindrakoto M, Roubaud F (2005) Governance, democracy and poverty reduction: lessons drawn from household surveys in sub-Saharan Africa and Latin America. Internal Metagora document.
- Jabine TB (1989) Statistics and human rights: some recent developments. *Proceedings of the 47th session of the International Statistical Institute, Book 1*, pp 463–464
- Jabine TB (1985) What can statisticians contribute to the protection and advancement of human rights? *Proceedings of the 45th session of the International Statistical Institute, Book 1*, pp 85–86
- Jabine TB, Claude RP, eds (1992) *Human rights and statistics: getting the record straight*. University of Pennsylvania Press, Philadelphia
- Jabine TB, Lynch JP, Spierer HF (1995) Statistics in an international human rights treaty report. *Proceedings of the section on government statistics of the American Statistical Association*, pp 77–86
- Kaufmann D, Kraay A, Mastruzzi M (2005) Governance matters IV: governance indicators for 1996–2004 (draft) (www.worldbank.org/wbi/governance/pdf/GovMatters_IV_main.pdf, accessed September 3, 2006)
- Spiere HF (1990). Violations of human rights: how many? *American Journal of Economics and Sociology* 49:199–210
- Spiere HF, Spiere L (1993) Data analysis for monitoring human rights. AAAS, Washington

Questionnaire for an Inventory of Initiatives Measuring Democracy, Human Rights, and Governance

The present questionnaire intends to identify any relevant initiatives and related indicators which contribute directly or indirectly to the measurement of democracy, human rights, and governance.

Your organization may not call or define its initiative(s) and indicator(s) as proper measurement tools of democracy, human rights, and/or governance. However, those initiatives may still collect information in a systematic way or document and measure some aspects or dimensions of democracy, human rights, and/or governance. For instance, indicators commonly denominated as social and economic indicators may be considered as a useful source of information in situational analysis of social and economic rights (e.g., adult literacy rate, daily per capita supply of calories, percent of infants with low birth weight). For this reason, we would like these types of initiatives also to be included in the inventory. As a consequence, questions in the questionnaire may not always apply to such an initiative carried out by your organization. Please feel free to leave them out.

The answers you provide will be used to categorize your initiative in a public database put on-line on the Internet. By answering as many questions as accurately as possible, you will make it easier for a future user of this inventory to locate your particular initiative, and thus get in touch with your organization.

Please feel free to add any additional information that you think would be useful for the understanding of your initiative. This information can come in the form of publications, project outlines, questionnaires etc. Such material would of course be only for internal study and will not be published without your prior consent.

If you know of other organizations in your country that have recently collected in a systematic way data or information on democracy, human rights, and governance, we would be grateful if you could send us their contact details (name of responsible person, e-mail, telephone or fax number) and forward our message or the address www.metagora.org/initiatives to them. All relevant documents can be downloaded from the link.

For any further questions, please do not hesitate to contact:

initiatives@metagora.org

Tel.: +33 1 45 24 91 15

Fax.: +33 1 44 30 62 78

A) Description of your organization

- | | |
|------------------------------------|---|
| 1. Name of the organization | Indicate the name of the organization(s) directly involved in the implementation of the initiative. |
| 2. Contact address | Indicate your organization's contact address:
Street:
ZIP/City:
Telephone:
e-Mail: |

3. Website of the organization	Indicate the organization's website. http://www.
4. Country	Indicate the country your organization is based in:
5. Type of organization	Choose one or more of the following: <input type="checkbox"/> Academic (University) <input type="checkbox"/> NGO <input type="checkbox"/> Governmental Organization <input type="checkbox"/> Intergovernmental Organization <input type="checkbox"/> National Statistical Institute <input type="checkbox"/> United Nations <input type="checkbox"/> National Human Rights Institution <input type="checkbox"/> Cultural organization <input type="checkbox"/> Research center <input type="checkbox"/> Other _____
6. Other organizations involved in the implementation of the initiative	If the initiative is implemented in collaboration with other organizations, please give their name and contact address:
7. Organization's main field of expertise	Indicate the field your organization is specialized in: <input type="checkbox"/> Statistics <input type="checkbox"/> Democracy <input type="checkbox"/> Human Rights <input type="checkbox"/> Governance <input type="checkbox"/> Other: _____

B) Initiative contributing to the measurement of Democracy, Human Rights, and Governance
 (If your organization is implementing or has implemented more than one such initiative, please fill in a form for each of them.)

a) General characteristics

8. Name of initiative	Indicate the specific name of the initiative.
9. Website of/on the initiative	Indicate the specific website for the initiative (if more specific than the website of organization); in case your initiative does not have an own webpage, you may indicate a website where information about it can be found. http://www.
10. Contact address	Indicate contact address for the initiative (if more specific than the contact address of the organization): Name (of responsible person or department within the organization): Street: ZIP/City: Country: Telephone: e-Mail:
11. Publications	Indicate where published results can be found, as well as the title and year of the publication.
12. Cost/Funding	Indicate the approximate total cost of implementation of the initiative and if possible the source of funding: Total cost (specify currency): Sources of funding:

(Continued)

B) Initiative contributing to the measurement of Democracy, Human Rights, and Governance

(If your organization is implementing or has implemented more than one such initiative, please fill in a form for each of them.)

a) General characteristics**8. Name of initiative**

Indicate the specific name of the initiative.

13. Status of the initiative

Indicate the status of your initiative, choosing *one* of the following:

Ended (add if possible the time span from first design to final publication: from ____ until ____)

Ongoing (since ____ until ____)

14. Geographical location

Indicate the geographical location (continent/region/country/area) your initiative is/was implemented in.

b) Objectives of the initiative**15. Purpose**

Give a brief description of your initiative (What are the general objectives? What was the need or demand that prompted the initiative? Are there links to specific laws, human rights treaties, or norms? Which is the target population? etc.)

Maximum length: 1500 characters.

16. Themes addressed

Check the circle(s) with the main theme(s) covered by the initiative on the left; if applicable, you may specify in the column on the right (Please select a maximum of six themes).

Democracy Popular sovereignty

Legitimacy of rulers Pluralism of political parties

Division of powers Rule of law

Elections Equality

Media freedom other: _____

Human rights Political rights

Civil rights Economic rights

Social rights Cultural rights

Environmental rights Right to development

Women's rights Children's rights

Rights of indigenous people and minorities

Other: _____

Governance Resource management

Accountability Transparency

Corruption Participation

Efficiency Rule of law

Control/Monitoring/Supervision

Access to information Ethics

Other: _____

c) Source(s) and production of data / information**17. Source(s) of data**

Indicate which source(s) of data are used, choosing *one* of the following:

Own source of data

Using other existing source(s) of data

Using a combination of own data and existing data

Initiative focusing on analysis not based on data

18. Type of data collection

Indicate type or method of data collection, choosing one or more of the following:

Census

- Random sample population survey
 Events registration
 Secondary sources (literature, newspapers, etc.)
 Administrative data (land records, etc.)
 Focus groups
 Panel of experts
 In-depth interviews
 Performance assessment / Desk studies
 Aggregation of multiple indicators using various data sources
 Other: _____
- 19. Specifications of type of data collection**
- You may wish to provide specifications or technical details about the type of data collection, such as population and sample size, nature of secondary sources or administrative data etc., level to which data can be disaggregated (country, village, household), etc.
- 20. Frequency of data collection / analysis**
- Indicate the frequency of data collection, choosing *one* of the following:
- Continuous (as for instance for events registration)
 Indicate time span: since _____ till _____
 Repeated (Indicate frequency of repetition (annual, every 5 years, etc.): _____)
 Unique
- 21. Measurement methods / tools**
- Describe the (quantitative and qualitative¹³) measurement methods and/or measurement tools generated or used.
 Maximum length: 1500 characters
- 22. List of indicators**
- Indicate the list of main indicators and possibly their definitions/descriptions.
- d) Outcomes and users**
- 23. Main outcomes (products)**
- Indicate which are the main outcomes/products generated by the initiative (e.g., reports, methods developed, database, etc.)
 Maximum length: 500 characters.
- 24. Main users**
- Indicate who are the main users of the information generated by the initiative, choosing *one or more* of the following:
- Policymakers Donor agencies
 International agencies Civil society
 Researchers Media
 Other: _____
-

Please return to:
 initiatives@metagora.org
 or

¹³ Information may be either qualitative, quantitative, or a combination of the two. Quantitative information has usually units of measurement, such as total number of persons, percent of population, percent of children under 15, people, years. Qualitative information does not have measurement units. For quantifying qualitative information, a scale would generally be required (e.g., an evaluation using a scale as 1 “totally agree,” 2 “partially agree,” 3 “neither agree nor disagree,” 4 “partially disagree,” and 5 “totally disagree”).

Metagora Co-ordination Team
OECD/PARIS 21
2, rue André Pascal
75775 Paris Cedex 16
France

or

Fax: +33 1 44 30 62 78

Reminder: If you happen to know other organizations in your country that are gathering data on democracy, human rights, and governance, we would appreciate if you sent us their contact address, and forwarded them the relevant documents or the address www.metagora.org/html/activities/act_inventory.html from where the relevant documents can be downloaded.

Chapter 9

Human Rights of Statisticians and Statistics of Human Rights: Early History of the American Statistical Association's Committee on Scientific Freedom and Human Rights

Thomas B. Jabine and Douglas A. Samuelson

9.1 Introduction

One warm early summer evening in 1986, Doug Samuelson's wife called him to their kitchen to take a phone call. When he answered, Samuelson got what he later described as "one of the nicest surprises, and overall highlights, of my entire life," as the caller explained, "This is Victor Kipnis. I'm calling to thank you. I'm in Toronto!"

Kipnis' release from the Soviet Union was the culmination of several years of effort on behalf of Soviet "refuseniks"—citizens who had been denied permission to emigrate—by the American Statistical Association's Committee on Scientific Freedom and Human Rights. Samuelson had delivered a paper by Kipnis, Eugene Grechanovsky, and Ilya Pinsker at the 1985 American Statistical Association (ASA) Annual Meetings to help dramatize their situation (Pinsker et al. 1985). The paper, smuggled out of the Soviet Union, represented work these statisticians had done in spite of pressure from the Soviet authorities. At the time of the presentation, Committee members invited those present to sign petitions to Soviet authorities on behalf of the authors. Samuelson also prepared a camera-ready copy of the paper for the proceedings volume.

Readers may find it unremarkable that ASA made such an effort to help Soviet statisticians and that it has a Committee on Scientific Freedom and Human Rights. Yet that committee is a relatively late arrival in ASA's 165-year history, and it was not established without opposition. In this chapter, the authors, both former chairs of the Committee¹, will describe how and why the Committee was formed in the late 1970s, and how its charge was expanded, again with some difficulty, to cover not only work on behalf of statisticians whose rights are believed to be violated, but also the encouragement of the use of sound statistical methods in monitoring the status of human rights everywhere. We will also deal briefly with attempts, which

¹ Jabine was chair from 1981 to 1984 and Samuelson from 1985 to 1988.

were less successful, to persuade the International Statistical Institute to engage in similar activities. Most of the events we describe took place between 1976 and 1988.

Except for a short paper by Jabine (1985a), documentation of the early days of the Committee is limited, and the authors have had to rely largely on memories of the past. We were aided in this task by reviewers of early drafts, and would like to thank Fred Leone, R. Clifton Bailey, Herbert and Louise Spierer, and Eric Stover for their valuable assistance.

9.2 A Victim of the “Dirty War”

In the mid-1970s, ASA’s Committee on International Relations in Statistics received a grant that enabled it to organize an exchange program with Latin American statistical societies. As part of this program, a delegation of ASA members visited Buenos Aires in May 1976. Carlos Noriega, who was then director of the Argentine national statistical office, was a member of the host committee. Later that year, Noriega left his post. Informal reports from colleagues indicated that he had been forced out because he refused requests from the newly established military government to tamper with official data series. He began a career as an independent professional consultant. He had a short-term assignment in Peru for the UN and UNICEF and was expecting to do additional work for the UN Statistical Office.

Early in February 1977, while on vacation in Mar del Plata with his wife and children, Noriega was detained by persons believed to be agents of the government or members of paramilitary groups. The government never acknowledged that he was in custody. Presumably, he was executed, one of the thousands of victims of Argentina’s “dirty war.”

9.3 Establishment of the Ad Hoc Committee

Fred Leone, who served as Executive Director of ASA from 1973 to 1988, had been instrumental in organizing the exchange program and was a member of the US delegation that visited Buenos Aires. When it became fairly obvious what had happened to Noriega, he was deeply concerned². In addition, at about that time, the International Statistical Institute (ISI), which would be meeting in December 1979³ in Manila, announced tentative arrangements to accept an invitation by the Argentine government to hold its 1981 session in Buenos Aires. In view of what had happened to Noriega, should there be an attempt to persuade the ISI to choose

² Fred Leone has informed us that the ASA delegation, while in Santiago, had learned that the Chilean government was secretly maintaining a database of “subversives” in the Southern Cone of Latin America and making it available to other countries wishing to take action against those whose names were listed.

³ The ISI meets biennially, in odd-numbered years.

a different site? If such an attempt failed, what official positions, if any, should ASA take about its members' participation in the 1981 session? Leone persuaded ASA's Board of Directors to establish an Ad Hoc Committee on Scientific Freedom and Human Rights with the mission of determining how ASA might respond to questions arising from abuses of the human rights of fellow professionals.

One source of support for efforts of that kind was the American Association for the Advancement of Science (AAAS). Several of AAAS's other constituent scientific societies were already active in advocating scientific freedom and human rights, and AAAS had a small staff collecting and disseminating information about scientists whose rights were being abused and assisting societies in preparing and coordinating their responses. R. Clifton Bailey, then serving as ASA's liaison to AAAS, joined in the effort and recommended two possible chairs for ASA's Ad Hoc Committee: Edward W. Barankin, of the University of California, Berkeley, and Dorothy Wellington, of the Environmental Protection Agency. Barankin, in particular, had been one of the first US statisticians to raise concerns about the ISI's planned meeting in Buenos Aires.

ASA's Ad Hoc Committee held its first meeting in August 1979 during ASA's annual meeting in Washington DC, and Ed Barankin was appointed chair. The Committee quickly decided that its first order of business was to propose an ASA response to the ISI's choice of meeting sites. Our recollection is that a majority of the Committee favored an organized boycott of the meeting in Buenos Aires if the situation in Argentina did not change and the ISI did not move the meeting, but no formal vote was taken at the time. Instead, the chair and ASA's Executive Director (Leone) were authorized to do their best to bring about a change in ISI's decision.

9.4 The 1981 International Statistical Institute Session in Buenos Aires

At the 1979 ISI session in Manila, during the organization's business meeting, Fred Leone and Ed Barankin argued unsuccessfully that the planned site of the 1981 session should be changed. Informed of this outcome, the Ad Hoc Committee considered various ways in which ASA might demonstrate its concern for what had happened to Carlos Noriega. Its first recommendation to the ASA Board of Directors was an official boycott of the ISI session. The Board rejected this proposal, with some members being so strongly opposed that they suggested dissolving and reconstituting the Committee. At one meeting, only a strategically timed lunch break, called by Fred Leone, kept this proposal from coming to a vote. Some members of the profession, learning of the controversy through the ASA monthly newsletter (*Amstat News*), wrote letters opposing the Committee's position, urging that science and politics should be kept separate. (Dorothy Wellington remarked dryly, "Now if someone would just convince these governments of that. . .")

Over the next several months, the Committee proposed several other measures to the Board of Directors, e.g., that the ASA would not pay for the travel of any of its members to the ISI session. This proposal was also rejected; however, by a

narrow margin, the Board approved a proposal that the Ad Hoc Committee organize activities before and during the session to show ASA members' concern about the status of Carlos Noriega and Graciela Mellibovsky, another "disappeared" Argentine statistician.

The Ad Hoc Committee was greatly aided in these activities by the Clearinghouse on Science and Human Rights⁴ of AAAS, and in particular by Bruce Kiernan and his successor Eric Stover, project directors for the Clearinghouse. Stover prepared a report, *Scientists and Human Rights in Argentina since 1976* (Stover 1981), which was included in an information packet that was made available to ASA members. The Ad Hoc Committee also drafted a petition, requesting information about the status of Noriega and Mellibovsky, to be presented to Argentine authorities at the time of the session, and circulated it to prospective participants in the session. The information packet and petition were publicized in *Amstat News*.

During the ISI session, interested ISI members met with Argentine human rights leaders (conveniently, the headquarters of a major human rights group was located across the street from the ISI meeting site) and families of disappeared scientists. A group of ISI officials met with the acting minister of interior and requested information on Noriega and Mellibovsky. Copies of the petition were delivered to Argentine authorities and to the newspapers. Surprisingly, the petition, along with the names of several signers, was published in a major Buenos Aires newspaper shortly after the session. Stover attended the session and helped to coordinate these activities. Dorothy Wellington was an active participant.

9.5 Efforts on Behalf of Other Statisticians

Regrettably, these efforts did not help Carlos Noriega or Graciela Mellibovsky. In all probability, Noriega had been executed well before the ISI meeting in Buenos Aires, and later efforts by the ASA's Ad Hoc Committee uncovered strong evidence that Mellibovsky had been executed shortly after she disappeared. But the Committee learned, from the AAAS Clearinghouse on Science and Human Rights and other sources, that there were statisticians in other countries who were apparent victims of human rights abuses. Such individuals included Abram Kagan and Yosif Begun, Soviet statisticians who were being refused permission to emigrate; Samuel Greene, a Liberian statistician who had been arrested in April 1980 following the Samuel Doe coup; and Wen-chen Chen, a professor of statistics at Carnegie-Mellon University and a Taiwanese citizen, who had died under suspicious circumstances during a visit to Taiwan in July 1981. The Committee decided that it would also be appropriate to extend its concern to scientists in disciplines closely related to statistics, one of whom was José Luís Massera, an internationally known Uruguayan mathematician who had been detained by the military government in 1975 and sentenced to 20 years imprisonment for alleged subversive activities.

⁴ Later renamed the Science and Human Rights Program.

Gradually, with the help of experienced human rights advocates from the AAAS Clearinghouse, the Ad Hoc Committee developed procedures for actions that could be taken on behalf of these individuals. Specific appeals to a government that it take some action affecting one of these persons required approval of the Board of Directors. Simple requests for information from government authorities about the status of an individual could be submitted by the Committee on its own initiative and, in cases calling for quick action, such inquiries could be initiated by the Committee chair and vice-chair, with the approval of ASA's executive director. We believe that such requests for information did, in some instances, contribute to improvements in the status of the individuals in question. In the case of Samuel Greene, in particular, the Committee eventually heard from Greene himself that he believed the Committee's letter to the Government of Liberia had saved him from execution.

9.6 From Ad Hoc to Continuing Status

With the expansion of "case work" and the identification of other human rights matters of potential interest to statisticians, it soon became clear that there was

Charge of the ASA Committee on Scientific Freedom and Human Rights—May 1982

The committee will concern itself with violations of and threats to the scientific freedom and human rights of statisticians and other scientists throughout the world. The following shall be its functions:

1. To gather factual information about current developments in its areas of concern. This function may include communications with appropriate authorities to inquire about specific cases that come to the committee's attention.
2. To inform the members of the American Statistical Association, through *Amstat News* and by other means, of cases that come to the attention of the committee, inviting response from the membership.
3. To encourage discussion of scientific freedom and human rights matters in meetings of the ASA, its sections and chapters, and other scientific societies, and in publications of these organizations.
4. When deemed appropriate, to recommend to the Board of Directors or the Executive Committee of ASA actions to defend or promote the scientific freedom and human rights of statisticians and other scientists.
5. As directed by the Board of Directors, to implement such actions.

Under no circumstances will the committee or any of its officers or members make any commitment or take any position on behalf of the association or its membership without prior authorization by the Board of Directors.

In carrying out these functions, the committee may avail itself of collaboration with similar groups in other scientific societies, in particular with the Committee on Scientific Freedom and Responsibility of the American Association for the Advancement of Science.

Fig. 9.1 Charge of the Committee on Scientific Freedom and Human Rights of the American Statistical Association, 1982

need for a continuing committee⁵ to work on these issues. Members of the Ad Hoc Committee developed a proposed charge for such a committee and, as required by ASA's by-laws, submitted it for approval to the Committee on Committees. That committee was not satisfied with the initial version, raising questions about how the proposed committee would relate to the association and its board of directors, and to what extent it would be making statements on their behalf. The proposed charge was revised, and the new version was approved by the Committee on Committees and subsequently by the Board in May 1982. The original charge of ASA's Committee on Scientific Freedom and Human Rights is shown in Fig. 9.1. The wording of the charge clearly reflects the concern of some of ASA's officers and directors that the Committee not become a "loose cannon," taking potentially controversial positions and actions without prior approval. The final paragraph includes an acknowledgment of the valuable assistance that the Committee had received (and would continue to receive) from the AAAS.

9.7 The Wen-chen Chen Case

The case of Wen-chen Chen brought the Committee's work to the attention of many statisticians for the first time. Morris de Groot, chairman of the Statistics Department at Carnegie Mellon University and a member of the Ad Hoc Committee, had flown to Taiwan shortly after Chen's death, accompanied by a forensic pathologist. Aided by the pathologist's investigation, de Groot was convinced that Chen had been thrown from the staircase where his body was found, rather than having fallen or jumped as the official account claimed. This question had practical importance for Chen's family, because Chen had a life insurance policy from a company based in Taiwan. Based on the official conclusion of "probable suicide," the company refused to pay the claim. Appeals from de Groot and other colleagues of Chen produced no response. Subsequent efforts to persuade the Taiwanese government and courts to open a formal investigation into Chen's death, or simply to rescind the official ruling of suicide, have so far been unsuccessful.

De Groot proposed that the Committee and ASA highlight the situation, in the hope of putting more pressure on the Government of Taiwan to change its position. In the fall of 1982, therefore, the Committee decided to sponsor a session at the 1983 Annual Meetings about Chen's work and what had happened to him. The program for the meetings had already been set, but Ed Bisgyer, the Associative Executive Director of ASA, pointed out that a committee could have a meeting whenever and wherever a room is available and could set and publish any agenda it chose. The Committee chose to announce publicly that its meeting would focus largely on Wen-chen Chen and invited his brother Robert to serve as the main speaker on the topic. The session attracted over 100 people and served as a vehicle for identifying additional ASA members who were interested in human rights issues.

⁵ Now referred to in ASA's by-laws as "current committees." The need for such committees must be reviewed at least once every 7 years by the Board of Directors.

9.8 Cases from the Soviet Union

The years 1982 and 1983 brought several additional Soviet statisticians to the Committee's attention, mostly via AAAS and the Committee of Concerned Scientists, whose human rights advocates were then focusing primarily on Soviet cases. In addition, in 1983, the Committee learned from AAAS that the Soviet Union had reportedly begun revoking academic degrees of Jewish scientists who had applied to emigrate from the country.

The reason for this action, as understood by those who reported it, was that the Soviet constitution guaranteed to all citizens employment appropriate to their education and experience. For those with advanced scientific degrees, virtually all of these jobs involved national security and therefore required clearance. Application to leave the country raised concerns about the applicant's loyalty and hence about the desirability of issuing or continuing clearance. Revoking the academic degree, rendering the individual ineligible for the job, apparently was considered to be a good way to handle the security risk while complying with the constitution.

Since the rationale appeared to be legalistic, the Committee settled on a response emphasizing legal concerns as well as the Soviets' desire to continue scientific exchanges with the West. The letter to Soviet authorities that was drafted by the Committee and approved by the Board pointed out that, for scientific exchanges to continue, it was necessary for US participants to be sure that their counterparts identified as scientists really are what they claim to be. If degrees can be revoked for political reasons, the letter continued, how are we to know that they are not also granted for political reasons? A number of other scientific societies wrote similar letters. While there was no formal response from the Soviet government, AAAS reported a few months later that the revocation of degrees had apparently ceased.

Over the next few years, the Committee took on additional cases from the Soviet Union and expanded its efforts on their behalf. Such efforts included the presentation of Pinsker et al. (1985) and the presentation of two papers, Grechanovsky and Pinsker (1987) and Resnikoff (1987), at the ASA Annual Meetings. Samuelson recalls that preparing the Resnikoff paper for the 1987 proceedings volume was a particular challenge, because "my computer didn't have Symbol font and the guy had used two-thirds of the Greek alphabet in his notation. I had to draw in all those η s and ζ s and φ s and ψ s by hand. Still, I knew I had the easy part of the task."

9.9 Organizing Sessions for Annual Meetings

Encouraged by the response to the Wen-chen Chen meeting in 1983, the Committee decided to do more to make other statisticians aware of human rights issues, both professional and technical. In 1984, at the ASA Annual Meetings the Committee co-organized and co-sponsored with the Social Statistics Section an invited paper session on social indicators, which included two papers on human rights monitoring. At the 1985 ASA Annual Meetings, the Committee and the Social Statistics

Section presented another, similar session on social indicators, including human rights monitoring; in addition, the Committee received approval for another invited paper session, sponsored solely by the Committee, featuring one presentation on arms control and another on human rights issues. When Ed Barankin passed away in the spring of 1985, the Committee dedicated this latter session to his memory.

9.10 Using Statistics to Monitor the Status of Human Rights

Other chapters in this book make it abundantly clear that statisticians, in addition to acting on behalf of colleagues whose rights are violated, can apply their professional knowledge to the collection and analysis of quantitative data about the status of human rights in their own and other countries. International human rights treaties cover not only civil and political rights, but also economic and social rights relating, e.g., to education, nutrition, and health care, topics to which our profession has long given its attention. However, awareness of this possible expansion of the Committee's activities came only gradually to its members and, when it did, their proposal to add new activities to the Committee's charge was not readily accepted.

Late in 1981, in response to reports that Salvadorian citizens were being tortured and murdered by elements of the country's armed forces, the US Congress passed an amendment to the Foreign Assistance Act of 1961 making continued foreign aid to El Salvador contingent on progress by that country in bringing an end to these human rights violations. A report to the Congress was required by January 1982, and subsequent legislation required periodic certifications of progress. To meet these requirements, the State Department established a data collection system based largely on press and embassy reports. Human rights NGOs established their own data series, which showed substantially higher numbers of killings and other violations. In 1983, in his ASA presidential address, R. L. Anderson noted this politically controversial "numbers game" and suggested that these reports would be a suitable subject for statistical review and inquiry (Anderson 1984). As the controversy continued and grew, members of the Committee concluded that its mission should be expanded to allow statisticians, as Richard Savage later put it in his 1984 ASA presidential address, "... to work impartially to allow the decision-making process to be more enlightened" (Savage 1985).

Early in 1983, the Committee drafted a proposal to expand its charge to allow it to provide information and assistance to scientific societies and other organizations on statistical questions relating to the measurement, evaluation, and analysis of data on human rights. Although the proposal was given a sympathetic hearing by the Board of Directors, it was defeated by a narrow margin, because some members feared that the committee's new activities would cause ASA to get involved in controversial partisan political issues, such as the use of statistical data to determine the fate of aid to El Salvador.

In 1984, ASA President I. Richard Savage devoted his presidential address (Savage 1985) to what he called "hard/soft problems," by which he meant problems that are both difficult to solve (hard) and difficult to define precisely (soft.) He gave

four major examples, one of which was human rights monitoring. He strongly urged that the entire profession focus more attention on problems of this type, especially the four he had identified. Encouraged by President Savage's advice, Committee members resolved to try again to expand the charge. Discussion with key members of the Committee on Committees led to suggested language that limited the scope of the new activities to responding to requests from other scientists and scientific organizations on statistical questions relating to human rights data. The new draft of the Committee's charge was accepted by the Committee on Committees in August 1985 and by the Board in October 1985. The following were the key provisions added at that time:

1. Exchange information with other scientists and scientific organizations on
 - the scientific freedom and human rights of statisticians and other scientists generally;
 - the situation regarding specific individuals or groups; and
 - statistical questions relating to data on human rights.
2. Assist scientific societies or other responsible organizations, upon request, in statistical questions relating to the measurement, evaluation, and analysis of data on human rights.

9.11 The Book Project

In 1985, the AAAS received a grant from the Ford Foundation for a research project to evaluate the role of statistics in the documentation of human rights violations. Eric Stover served as project director. Jabine was asked to serve as chair of the project steering committee and Samuelson as one of its members. Agreement was reached with the editor of *Human Rights Quarterly* to devote a special issue of that journal to statistical issues in the field of human rights. The steering committee circulated a call for papers to several statistics and political science newsletters and to individuals known to be doing relevant work. Six of the proposals submitted were chosen by the steering committee for commissioned papers, and two additional papers were accepted for publication in the special issue (Claude and Jabine 1986). Updated versions of most of these papers and several additional papers on related subjects were later published in *Human Rights and Statistics: Getting the Record Straight* (Jabine and Claude 1992).

Although the ASA Committee was not a formal sponsor of this project, several of its members were actively involved. Herbert Spirer and David Banks, both of whom later served with distinction as chairs of the Committee, contributed papers. Spirer and Samuelson coauthored a chapter, "Use of Incomplete and Distorted" "Data in Inference About Human Rights Violations" (Spirer and Samuelson 1992) and Banks wrote a chapter, "New Patterns of Oppression: An Updated Analysis of Human Rights" "Data" (Banks 1992).

9.12 Human Rights and the International Statistical Institute

In 1982, a group of European statisticians established the European Working Group on Statisticians and Human Rights to work on behalf of some of the same individuals who were receiving the support of the ASA Committee on Scientific Freedom and Human Rights and its predecessor Ad Hoc Committee. During the 1983 ISI meeting in Madrid, the European Working Group held an informal meeting, which was attended by some members of the ASA Committee who were also members of ISI. Spurred by this initial contact, they began to consider ways in which they might stimulate ISI to become more involved in human rights matters.

It seemed to them that it would be reasonable for ISI to include, in its scientific meetings and journal, papers making use of statistics to monitor human rights. Looking at this question in a broad context, it was clear that several papers from prior biennial sessions of ISI had dealt with statistics of education, health, income, and other topics covered by international treaties on economic and social human rights. Asking ISI to support individual statisticians whose rights were being violated was clearly a more sensitive issue. Actions implying criticism of specific governments might interfere with ISI's goal of widespread participation in its scientific activities by statisticians from all countries.

In preparation for the 1985 ISI session in Amsterdam, three initiatives were planned:

1. Organization of a contributed paper session on "Statistics, Statisticians and Human Rights."
2. Preparation of a petition to Soviet authorities on behalf of Yosif Begun, a Soviet refusnik, to be circulated to ISI members for their signatures.
3. Preparation of a petition to the ISI council to review ISI's policies with respect to human rights activities.

As described in the following paragraphs, these initiatives were the source of some controversy and met with only limited success.

Tom Jabine, jointly with Jean-Louis Bodin of France⁶ and Ann Mitchell⁷ of the United Kingdom, submitted a proposal for a topical contributed paper meeting on "Statistics, Statisticians and Human Rights." Initially, the proposal was accepted and the meeting was listed in Bulletin Number 1 for the ISI session. However, the ISI Bureau subsequently decided that the topic was not appropriate and announced in Bulletin Number 2 that the proposed meeting had been eliminated from the program. Nevertheless, all ISI members are entitled to present a contributed paper at each ISI session, so Jabine and Professor Helmut Muhsam of Israel submitted papers on topics related to human rights and requested that they be scheduled for consecutive presentation at the same contributed paper meeting.

Perhaps by chance, it turned out that the two papers were scheduled for presentation on the same morning as an informal open meeting on human rights and

⁶ He later served as President of the ISI.

⁷ She served as Secretary of the European Working Group on Statisticians and Human Rights.

statistics, which had been requested by several members. The conflict was resolved by convening the open meeting, adjourning it temporarily so that those present could go to the adjoining room to hear the two contributed papers, and then proceeding with the informal meeting. As is customary, the two papers were published in the ISI Bulletin (Jabine 1985b; Muhsam 1985).

A draft of the Begun petition was approved unanimously at the informal meeting, signed by those present, and posted on a bulletin board for additional signatures. When the president of ISI saw the petition, he objected strongly to the fact that ISI was mentioned by name in its opening sentence. Although that sentence was an accurate statement of the circumstances under which the petition was developed, he was afraid that Soviet authorities receiving the petition might not understand that the informal meeting was not part of the official ISI program. After protracted negotiations, the sponsors of the petition (Bodin, Mitchell, and Jabine) agreed to change the first sentence and were promised that the ISI secretariat would send a letter from the sponsors to those who had signed the original version, explaining the change and inviting them to sign the new version. The new version was also posted on the bulletin board, and all signed copies were transmitted to Ann Mitchell, who forwarded them to the Soviet authorities.

A draft version of the petition to establish a committee to review ISI's policies on human rights matters was discussed at the informal meeting and some changes were proposed. A revised version was given to persons who had attended the meeting and had indicated their willingness to circulate the petition to obtain signatures. The signed versions were turned in to Bodin, who presented them to the ISI council, of which he was a member. No formal report of action on this proposal was ever issued by ISI.

9.13 Concluding Remarks

The adoption of the Universal Declaration of Human Rights in 1948, followed by the 1966 Covenant on Economic, Social and Cultural Rights and the 1966 Covenant on Civil and Political Rights, signaled the beginning of international human rights law-making by international and regional organizations, as well as a populist worldwide movement to ensure greater realization of the rights set forth in these treaties. Nongovernmental organizations, such as Amnesty International (founded in the mid-1960s), became an important part of this movement. National and international scientific and professional organizations gradually began to pay attention to human rights issues within the context of their primary interests.

ASA was neither the first nor the last professional association to take an active interest in these issues. For ASA, the decisive factors were the apparent violations of the rights of Carlos Noriega, the former head of a national statistical agency, and the determination of ASA's executive director, Fred Leone, that those violations should not go unnoticed. The work of the Ad Hoc Committee formed at his suggestion was greatly aided by assistance from the AAAS Clearinghouse on Science and Human Rights and by support from several ASA presidents during its early years.

Initially, the Committee concerned itself with assistance to statisticians and mathematicians, mostly in the Soviet Union, whose civil and political rights were apparently being violated. However, it soon became evident to members of the Committee that statisticians could also use their professional expertise to assist in the development of statistical indicators of human rights performance. After some early resistance, the committee's charter was amended to include such activities.

As we have explained, efforts to engage the International Statistical Institute in similar activities met with little success. ISI officers and staff were clearly apprehensive about activities that might imply criticism of member nations. Many ISI members are official statisticians whose participation in the organization depends on support by their national governments. There have been few instances of international professional societies openly criticizing a country or its citizens for human rights violations, one notable exception being the World Psychiatric Association's 1977 censure of Soviet psychiatrists for their participation in brutal treatment regimes for dissidents (Dean 1989).

We will leave it to others to carry this history forward from the end of Samuelson's tenure as chair of the ASA Committee on Scientific Freedom and Human Rights in 1988. Notwithstanding our lack of success in persuading ISI to become more involved, we feel that at this point the stage had been set for many new initiatives by members of the Committee in collaboration with others, including especially the AAAS Science and Human Rights Program and Human Rights Information and Documentation Systems International (HURIDOCS). The work of Herbert and Louise Spierer in training human rights advocates from all over the world on the use of statistics in monitoring human rights deserves a separate account. Their training manual, *Data Analysis for Monitoring Human Rights* (Spierer and Spierer 1993) has been translated into several languages.

One of ASA's goals, as stated in its constitution, is "... to increase the contribution of statistics to human welfare." In this spirit, as evidenced by the other chapters in this book, many statisticians, including ASA's immediate past president, Fritz Scheuren, are using their professional skills and knowledge to contribute to the advancement of the principles set forth in the Universal Declaration of Human Rights.

References

- Anderson RL (1984) Goals: where are we and where should we be going? Presidential address. *Journal of the American Statistical Association* 79:253–258
- Banks D (1992) New patterns of oppression: an updated analysis of human rights data. In: Jabine TB and Claude RP (eds) *Human rights and statistics: getting the record straight*. University of Pennsylvania Press, Philadelphia, pp 364–391
- Claude RP, Jabine TB (eds) (1986) *Symposium: statistical issues in the field of human rights*. *Human Rights Quarterly* 8:551–699
- Dean RN (1989) Nongovernmental organizations: the foundation of western support for the human rights movement in the Soviet Union. In: Claude RP and Weston BH (eds) *Human rights in the world community: issues and action*. University of Pennsylvania Press, Philadelphia, pp 299–312

- Grechanovsky E, Pinsky ISH (1987) Calculation of conditional p -values in a forward selection procedure. Proceedings of the Section on Statistical Computing of the American Statistical Association, American Statistical Association, Alexandria, pp 402–405
- Jabine TB (1985a) Human rights and scientific societies: Working from within. In: McCleskey K (ed) Scientists and human rights: present and future directions, Publication 85-19. American Association for the Advancement of Science, Washington, pp 39–46
- Jabine TB (1985b) What can statisticians contribute to the protection and advancement of human rights? 45th Session of the International Statistical Institute, Contributed Papers, Vol. I. International Statistical Institute, Amsterdam, pp 85–86
- Jabine TB, Claude RP (eds) (1992) Human rights and statistics: getting the record straight. University of Pennsylvania Press, Philadelphia
- Muhsam HV (1985) On two human rights involved in the ISI declaration on professional ethics. 45th Session of the International Statistical Institute, Contributed Papers, Vol. I. International Statistical Institute, Amsterdam, pp 475–476
- Savage IR (1985) Hard-soft problems. Presidential address. Journal of the American Statistical Association 80:1–7
- Pinsky ISH, Kipnis V, Grechanovsky E (1985) The use of the F-statistic in the forward selection regression algorithm. Proceedings of the Statistical Computing Section of the American Statistical Association, American Statistical Association, Alexandria, pp 419–423
- Spirer HF, Samuelson D (1992) Use of incomplete and distorted data in inference about human rights violations. In: Jabine TB and Claude RP (eds) Human rights and statistics: getting the record straight. University of Pennsylvania Press, Philadelphia, pp 62–80
- Spirer HF, Spirer L (1993) Data analysis for monitoring human rights. American Association for the Advancement of Science, Washington
- Stover E (1981) Scientists and human rights in Argentina since 1976, Publication 819. American Association for the Advancement of Science, Washington
- Resnikoff I (1987) Estimation of a number of outbursts for a stationary stochastic process. Proceedings of the Section on Statistical Computing of the American Statistical Association, American Statistical Association, Alexandria, pp 467–471

Chapter 10

Obtaining Evidence for the International Criminal Court Using Data and Quantitative Analysis

Herbert F. Spirer and William Seltzer

Abstract The judicial systems of many countries recognize the value of scientific evidence, including statistical analysis of data in many civil and criminal cases. Quantitative analyses of human rights violations can also have evidentiary value comparable to DNA testing, forensics, and chemical analysis. In this chapter, we describe and discuss evidentiary issues that arise in International Criminal Tribunals that try alleged perpetrators for serious human rights crimes. Our goal is to promote the effective use of statistical and demographic data and methods in these settings. Toward this end, we spell out their potential advantages in this particular legal environment and discuss problems in their use with the hope of stimulating open discussion among various actors.

10.1 Introduction

Our goal in this chapter is to identify and discuss the issues that relate to the effective use of statistics and demography in the International Criminal Court (ICC). This discussion draws on national experience in the use of data and quantitative analysis in legal settings, the limited but growing experience in the use of statistics and demography in the work of international criminal tribunals, and the special needs and circumstances that arise in the use of data and quantitative analysis in investigating, documenting, and analyzing serious human rights abuses.

Although the primary focus of this chapter is applications in the context of the ICC, most of the substantive issues addressed are also relevant to special international tribunals established in connection with country- or region-specific human rights crimes or the prosecution of serious human rights crimes under national laws. (For example, several persons have been convicted in Belgium courts for their actions in the 1994 genocide in Rwanda (“Belgium Finds Two Guilty in Rwanda Genocide Case,” downloaded from www.nytimes.com/reuters on 6/29/2005) and in the United States genocide is a federal crime when performed by US nationals or in the United States (18 USC Chapter 50A).) In addition, the issues discussed here may be of relevance to those involved with the work of truth commissions. Although we recognize that the methods and a number of the goals of truth commissions and

criminal prosecutions under international and national law differ, quantitative data and analysis can serve similar functions in both settings.

Before describing further the issues that we will address, it will be useful to review briefly the establishment and terms of reference of the ICC. The ICC as an institution arose out of lengthy international negotiations carried out under the framework of the UN with a view to establishing a permanent and continuing international criminal court. The diplomatic steps involved included a number of preparatory meetings and are reflected in a series of General Assembly resolutions (United Nations 1992, 1993, 1994b, 1995, 1996, 1997). Those negotiations culminated in a diplomatic conference of plenipotentiaries in the summer of 1998 in Rome, which by the adoption of its Final Act on in July 1998 set out the international agreement on the creation of the ICC and its mandate (CICC 1996–1998). The Treaty of Rome establishing the ICC came into force in July 2002 (ICC 2002). There is continuing public discussion related to the organization and work of the ICC (see, e.g., Louis-Jacques (2003) and CICC (2002–2005), as well as the ICC's own website, www.icc-cpi.int/home.html).

The Preamble of the Statute of Rome is helpful in defining the goal of the ICC when it affirms that

the most serious crimes of concern to the international community as a whole must not go unpunished and that their effective prosecution must be ensured by taking measures at the national level and by enhancing international cooperation.

Under article 5 of the Statute of Rome, the jurisdiction of the ICC is

limited to the most serious crimes of concern to the international community as a whole [which are specified as] (a) The crime of genocide, (b) Crimes against humanity, (c) War crimes, [and] (d) The crime of aggression.

It is clear that these crimes are among the core concerns of the human rights field. Although some international disagreement about the ICC remains (primarily the current objections to the ICC by the United States), as of mid-2006 it has begun investigations into three potential cases either at the request of individual member states, its own initiative, or on referral from the UN Security Council.

In carrying out these and other investigations and in any subsequent prosecutions and trials, the ICC is to be guided by the rules of evidence set forth in the Statute of Rome (principally article 69) and the Rules of Procedure and Evidence adopted for the ICC (rules 63 to 75) (ICC, Assembly of States Parties, 2002). While no explicit reference is made to statistics and data or even to expert witnesses, article 69, paragraph 3, does provide that the ICC “shall have the authority to request the submission of all evidence that it considers necessary for the determination of the truth.” Along the same lines, the Statute of Rome, in article 54(1), places a direct responsibility on the Office of the Prosecutor to “establish the truth” and to “take appropriate measures to ensure ... [an] effective investigation.” Toward these ends, the ICC has engaged the expertise of the Netherlands Forensic Institute as part of its pre-trial investigations into alleged human rights abuses in the Republic of the Congo (ICC 2005).

Moreover, paragraph 6 of article 69 provides that “The Court shall not require proof of facts of common knowledge but may take judicial notice of them.”¹ In many national jurisdictions, population census data and other outputs of the national statistical system fall into this category and courts may simply take judicial note of them.

Four attributes of the needs and circumstances of the ICC must be taken into account in considering any statistical or demographic activities likely to be carried out under its auspices:

1. The criminal events under investigation usually accompany prolonged and pervasive social chaos and often, massive physical destruction. As a consequence, many normal data sources are interrupted or negatively impacted by the crimes under investigation or by the military and political activities that end them. Crimes against groups of people and continuing population movements disrupt family units, which are often the key element in population data collection.
2. The physical and human conditions during investigations of crimes against humanity create severe time pressures. Much of the evidence and its sources—memory, victims, survivors, damaged sites, and physical records—are perishable. Often, the investigators themselves are under physical duress, hostile observation, or receive overt threats. In addition, national and international pressures for immediate action can be intense, often because of the urgent need to construct a civil society.
3. The ad hoc and international character of the process of investigation reduces efficiency. Investigators from different cultures and legal traditions must devote intense time and effort on short notice. They must “learn the ropes” of the affected region and its several component parts and the international and national rules governing the total effort, and find ways to operate in a multi-cultural and interdisciplinary environment. The ICC could institutionalize some of this knowledge through organizational learning by permanent staff, but the need to set up ad hoc investigative operations in different settings and circumstances will remain.
4. Sufficient funds must be made available. Funds made available for the previous international tribunals were always limited and the amounts initially allocated to investigation work disproportionately low.

To date, the ICC has had limited involvement with quantitative data, although we assume that the ICC will in time, like most courts, make considerable use of statisticians and others who use statistical and demographic methods to assist the court in its task of determining the truth. The range of evidence that will be used in connection with the “most serious crimes” under the ICC’s jurisdiction is broad.²

¹ The legal concept of “judicial notice” permits parties to introduce as evidence certain generally accepted factual matters without having to establish their factual character (e.g., tide tables or the time of sun rise or sun set). The ICC statute simply accepts the concept of judicial notice.

² See articles 6, 7, and 8 of the Statute of Rome, which specify in detail the acts that comprise genocide, crimes against humanity, and war crimes. A definition of the crime of aggression is still being developed.

However, in this chapter we are concerned with a subset of this evidence, that which uses quantitative data and demographic and statistical analysis. Many other types of evidence, both testamentary and scientific, enter into the deliberations of courts. These types of evidence are outside the scope of this chapter.

When we refer to “quantitative data and demographic and statistical analysis” in this chapter, we mean the body of techniques comprising the collection, compilation, analysis, and presentation of data. This body of techniques is often subsumed under the disciplines of statistics and demography. The data sources with which we are specifically concerned include national and local statistical and administrative systems, refugee, human rights, or humanitarian organizations, or the work of a court’s investigators. These data may describe persons (who were alive before the events or survivors, victims, those forced to move, missing persons, witnesses, etc.), forensic evidence (from graves), reports of crimes, or other evidence to be summarized or quantitatively analyzed.

There are many issues of international relations, jurisdiction and jurisprudence, and public policy to be addressed as part of the work of the ICC. The resolution of these issues is outside the scope of this chapter, but we acknowledge that these more general evidentiary issues may influence the use of data and quantitative methods.

The next section of the chapter first briefly reviews experience in the use of data and quantitative analysis, primarily statistical analysis, in the courts in the United States and in other jurisdictions. It then moves to a more extended discussion of issues that arise in using data and quantitative analysis in the investigation and prosecution of major human rights abuses. The third section reviews the experience to date in the use of statistical and demographic data and methods in the Nuremberg Tribunal, the International Criminal Tribunal for Rwanda (ICTR), and the International Criminal Tribunal for the former Yugoslavia (ICTY). The final section of the chapter briefly discusses some of the issues raised and presents our main conclusions.

10.2 The Role of Statistics and Demography

There are three main ways in which statistical and demographic methods and data can assist in the work of the ICC or other national or international criminal courts or tribunals:

1. Assisting in the investigative process.
2. Producing statistical or demographic estimates to be offered in evidence, either in the form of descriptive statistics, causal analysis, or other types of analysis.
3. Enhancing the format and presentation of data offered in evidence.

10.2.1 Relevant Prior Work at the National Level

All three of these methods have been used in several national jurisdictions. For example, in the United States there have been two recent efforts related to the

demonstration of probative use of statistics in the courts. The first, the Panel on Statistical Assessments as Evidence in the Courts (a joint effort of the National Academy of Sciences' Committee on National Statistics and its Committee on Research and Law Enforcement and the Administration of Justice) deals with

a variety of issues that arise in . . . court proceedings when statistical assessments such as quantitative descriptions, causal inferences, and predictions of events based on earlier occurrences are presented as evidence. [The panel appraises] the forms in which such assessments are presented, aspects of their admission into evidence, and the response to and evaluation of them by judges and juries (Fienberg 1989, p. xi).

The Panel also observed

statistical knowledge will continue to be needed and used in litigation. This use will include quantitative descriptions, causal inferences, and predictions of future events based on earlier occurrences (Fienberg 1989, p. 1).

The Panel noted that such usage by the courts had increased markedly in recent years and that it was expected to increase.

This Panel's approach is also relevant to the extension of the use of data and quantitative methods to the ICC. The panel participants pooled legal and scientific experiences, searched for relevant cases, listened to expert presentations, discussed the ethical and professional obligations of both disciplines, reviewed past litigation, and examined the literature. One tangible result of this panel's work was the publication of a report suitable for use as a reference manual (Fienberg 1989).

The second project, a joint effort of the U.S. Federal Judiciary Center and the Carnegie Corporation of New York, resulted in the publication of the *Reference Manual on Scientific Evidence: For Use with Federal Practice and Procedure* (Federal Judiciary Center 1994). In his review of this manual in *The American Statistician*, Joseph L. Gastwirth (1997, p. 297) said, "it is essential to ensure that trial judges have the scientific background and quantitative skills to make fair and uniform rulings about scientific testimony," and felt that this publication would be useful in achieving that goal. We believe that the ICC can benefit from activities and publications similar to these.

Although both projects concern only US practice, they show the issues that arise in courtroom application of data and quantitative methods. Moreover, statistical and other quantitative methods are also used in France (Besson 1992, pp. 66–76), Australia, New Zealand, Canada, Austria, Germany, and the United Kingdom (Robertson and Vigneaux 1995, pp. 3–4).

10.2.2 The Role of Statistical and Demographic Data and Methods at the ICC

In this section, we discuss in detail issues that arise in using demographic and statistical data in the investigation and prosecution of such major human rights abuses as genocide, crimes against humanity, war crimes, and aggression. This discussion includes evidentiary issues, basic data sources, the role of sampling, basic methods of analysis, processing and compilation issues, ethical considerations,

training needs, the costs of such work, standards for assessing statistical evidence, and the need to foster cooperation among lawyers, investigators, and scientists engaged in the work of the ICC.

Table 10.1 summarizes some of the major types of data, statistical and demographic methods, analytical applications, and methods of data presentation that are relevant to the work of the ICC. This table serves as a reference point for the more detailed discussion of these and related issues in the balance of this section.

10.2.2.1 Basic Evidentiary Issues

“How many?” is often the first question asked when dealing with evidence of genocide, crimes against humanity, or war crimes. Although statements such as, “a great many” or “thousands” are rarely very helpful, more precise numerical answers to this question can have legal value. For example, the definition of genocide includes the phrase, “acts committed with intent to destroy, in whole or in part, a national, ethnical, racial or religious group,” while the definition of crimes against humanity refers to “acts when committed as part of a widespread or systematic attack directed against any civilian population” (ICC 2002, articles 6 and 7).

In the jurisprudence of the International Tribunal widely adopted in the ICTY, “the phrase “widespread” refers to the large-scale nature of the attack and the number of targeted persons, while the phrase “systematic” refers to the organized nature of the acts of violence and the improbability of their random occurrence. Patterns of crimes, in the sense of the non-accidental repetition of similar criminal conduct on a regular basis, are a common expression of such systematic occurrence. Only the attack, not the individual acts of the accused, must be widespread or systematic.” (See, e.g., the ICTY, *Blaskic Appeal Judgement*, paragraph 101 (Case No. IT-95-14-A).)

Numbers indicating magnitude are also important with respect to the humanitarian law concept of proportionality, particularly relevant to the consideration of war crimes. The doctrine of proportionality asserts that weapons and military actions must discriminate between combatants and civilians (Roberts and Williams 1995, pp. 3–4). Military analysts hold that a certain ratio (“proportion”) of civilian to military casualties in military operations is unavoidable. How do we determine body counts to evaluate claims related to whether “all or part of a group were killed,” the attacks were “widespread or systematic,” or a “disproportionate” number of civilians were killed, in any given case?

Investigators and prosecutors of major human rights abuses (genocides, crimes against humanity, war crimes, and aggression) seek counts for violations such as deaths, disappearances, summary executions, torture victims (by type), sexual assaults (by type or by perpetrator), refugees (by origin, departure routes, destinations), forced removals, enslavement, maiming, degradation, evictions, and so forth. The numerical magnitude of violations of humanitarian law plays a role in creating pressure for the search for truth and the indictment and conviction of perpetrators. Ball and his colleagues examined this problem, saying

Table 10.1 Typology of data, methods, analytical applications, and methods of data presentation relevant to the work of the International Criminal Court

Macro-data on persons and violations	<ul style="list-style-type: none"> ● Estimates of the numbers of victims, the number of violations (murder, rape, etc.) for country, geographical region, perpetrator, ethnic identity, gender, age, affiliation, or administrative entity ● Estimates of the patterns of victimization and violation, by type of violation, over time, within different sectors of the population, gender, affiliation, or age ● Estimates of the extent and economic costs of property destruction and contamination or denial of access to property (e.g., as in the case of land mining)
Micro-data on persons and violations	<ul style="list-style-type: none"> ● Names of victims based on official lists of persons alive prior to a given date ● Names of victims and types of violations based on records of persons needing assistance during and after the period of alleged criminal activity ● List of persons alive after the period of alleged criminal activity
Methods of data collection and analysis	<ul style="list-style-type: none"> ● Enumeration of destroyed and contaminated property ● Use of prior censuses, sample surveys, or administrative registers to identify or estimate population at risk to inventory or value property at risk of destruction or damage ● Use of subsequent censuses, sample surveys, or administrative registers to identify or estimate survivors or to inventory or value remaining property ● Use of forensic studies ● Use of case studies ● Use of demographic analysis drawing on available data sets to (a) adjust data for likely deficiencies and (b) to estimate mortality or migration attributable to the alleged crimes under alternative assumptions ● Use of case-by-case matching of reports of persons or events from two or more sources to estimate number of persons or events adjusted for omissions ● Use of techniques of statistical analysis to test the plausibility of alternative explanatory hypotheses in light of the observed data
Methods of data presentation	<ul style="list-style-type: none"> ● Presentation of data (tables, charts, maps) to communicate effectively in the courtroom ● Integration analytical graphical presentations to reveal the structure and patterns in the data, such as geographical relationships, sources and time movements of violations, relationship of violation clusters to perpetrators, etc.

Can these counts be obtained? In some cases, investigators and statisticians can get accurate counts or estimates whose precision is measurable. For example, if there was a massacre in an isolated area with no removal of cadavers, with no degradation of the site or remains, then skilled investigators can get credible counts, as in Iraq, Russia (Babi Yar), El Salvador (El Mozote) and Mozambique (Frelick 1988). Three organizations investigating

of the state violence in Guatemala during the period 1960-1996 used media reports, field interviews, and documentary reports of non-governmental organizations to establish credible estimates (Ball et al. 1999, p. 11).

Demographers have techniques to test the quality of data taken from censuses, surveys, and administrative reporting systems. They can also make indirect estimates of mortality where no direct data exist and make credible estimates of mortality due to crimes against humanity (e.g., DaPonte 1993, pp. 57–66). The United Nations has produced manuals, handbooks, and texts giving guidance for making basic demographic estimates and provides software that supports these techniques (United Nations 1983a,b, 1990). These methods can be extended to violations other than killing, but there is a need for more theory, application methodology, and software for making credible estimates of mortality and the extent of other violations. The search for the answer to “How many?” sometimes leads to an unnecessary quest for an unachievable exactitude. For legal purposes it may be sufficient to set a scientifically supportable lower bound for the count, or to state a reasonable interval estimate. Given the difficulties of measurement and estimation, this prudent approach was used by the Guatemalan Commission for Historical Clarification (CEH 1999, Annexo III). Investigators can improve their results and credibility by consulting statisticians and demographers to verify their methodology both before and after data collection. For example, no statistician would support the use of a total count of the dead that was based on counts of the dead from just three burial sites that is then extrapolated to a total count based on the number of known sites. Yet such estimates are sometimes put forward by human rights advocates.

What kinds of estimates are suitable? A statistician wants estimates that are both unbiased and precise. Among other characteristics, an unbiased estimate is one that is based on a sample that is representative of the population from which it is drawn. The use of a probability sample is one way to increase the likelihood of such representation. The primary reason for expending the time and resources needed for probability sampling is to reduce bias; the second most important reason is that, with data from a probability sample, a statistician can scientifically state the precision of estimates computed from data. The method known as capture–tag–recapture (Marks et al. 1974, Fienberg 1972, 1992, Fienberg et al. 1999) used to estimate the completeness of civil registration systems and census coverage can also give a statistically sound approach to the estimation of the number of dead and missing and those forced to flee because of large-scale human rights abuses.

Estimates of the aggregate economic value of property damage are needed to determine the costs to victims of property destruction targeted against particular groups. For example, this occurs when a perpetrator targets a particular religion and damages, destroys, or appropriates church property belonging to that religion’s institutions. Investigators can, in some cases, assess damage through a complete enumeration. This is a statistical issue, since such a census calls for definitions of categories of destruction and collection, analysis, and verification of data. Where complete enumeration is impossible a statistician can be enlisted to take either a probability or a judgment sample to make an estimate of damages to a known precision.

“Is there a systematic pattern?” is also a question of considerable evidentiary importance. As already indicated, one part of the definition of crimes against humanity refers to “systematic” attacks which has frequently defined in terms of “the improbability of their random occurrence.”

Patterns of violation can also help (1) to establish the nature of the crimes (genocide, violation of medical neutrality, summary execution, etc.) and (2) to identify alleged perpetrators. For example, investigators can look for a rise in violations when a particular individual or organized group is active in a region. They may also look for a disproportionate level of violations against a particular ethnic group, and temporal or special coincidence between well-documented events and mortality or migration or between provocative statements and subsequent events (Ball 2000, Brunborg et al. 2003, Tabeau et al. 2003, Verwimp 2003 and 2004). Jumping from such associations to cause and effect, without other evidence, is rarely justified whether one is dealing with human rights abuses or other issues. However, patterns of association can be used to support other evidence thereby raising or lowering the likelihood of culpability, to explore the plausibility of alternative hypotheses for the events observed, or to suggest promising lines of investigation.

“Does the pattern of the data themselves reveal evidence of purposeful manipulation or concealment?” Investigators can analyze incomplete and distorted data for evidence. There are many examples of useful inferences in similar circumstances, including mortality statistics, burial time series, health statistics, and economic data. Such examples exist for Argentina, The Republic of South Africa, USSR, Cambodia, and Nazi Germany, among others (Samuelson and Spierer 1992, pp. 62–77).

“What is the best way to present the data?” Presentation of data is a quantitative art and science. Edward H. Tufte (1997, p. 52) makes a convincing argument that, in the case of the Challenger disaster in which a rocket exploded and killed seven astronauts in 1986, “Had the correct scatterplot or data table been constructed for use in the decision making process, no one would have dared to risk the Challenger in such cold weather.”

Skillful visual representation can help the investigator in drawing inferences (as it would have in the Challenger launch decision) and plays an important role in the courtroom. Graphical representation is directly applicable in the analysis and presentation of information about crimes against humanity. Such information can include, e.g., refugee flows and their temporal and spatial relationship with the actions of alleged perpetrators. They can help in relating crimes to the actions of individuals who are at a high level in the chain of command.

Tabular presentation, as opposed to verbal or textual description, is often effective in presenting evidentiary summaries of data. Many textual summaries of data are confusing and incomplete. Often they can only be comprehended by parsing the data presented back into tabular form to determine the categories omitted and to verify the correctness of the summaries.

Our limited set of queries here is illustrative. Lawyers, investigators, statisticians, and others working on other cases would certainly have further questions that could be addressed by the analysis, presentation, and collection of pertinent quantitative data.

10.2.2.2 Sources of Data

Five major sources of data that can be used, individually or in combination, to provide quantitative evidence on major human rights abuses are as follows: (1) censuses, (2) sample surveys, (3) administrative reporting systems, (4) medical and anthropological forensic studies, and (5) case studies. We review each of these data sources in turn.

Censuses

The primary census data source is the national population census carried out in almost every country of the world, usually at 5- or 10-year intervals. A population census is a complete or partial enumeration supplemented by sample enumeration (to obtain more detailed data than can be collected in the full census and, rarely, to reach enumeration-resistant groups in the population). The purpose of such a population census is to develop aggregate statistics (so-called macro-data) on the number, location, and demographic and socioeconomic characteristics of the population. It is not intended to gain information about individuals for any purpose other than the development of aggregate data.

Population censuses can be important sources of benchmark data on the size, location, and characteristics of the victims of crimes resulting in genocide or mass migration. The census data are also needed if the investigator wishes to compute rates of victimization based on the number of persons at risk. In many countries with social divisions and minority group problems, the census gives data on the ethnic, tribal, linguistic, and religious groups of concern. Some countries do not recognize the existence of particular internal divisions in the population as a matter of policy, such as Kurds in Turkey or, until recently, “Indians” in Mexico. “Know the subject matter” should be the investigator’s motto.

In addition, in several countries (e.g., Germany and Rwanda) where genocides have occurred, the use of the kind of statistical classifications used to identify and target the victim population was subsequently barred. While this clearly serves the policy use of discouraging such humanitarian crimes in the future, it may hamper the development of accurate estimates of the toll of the previous genocide.

The time gap between events of interest and the census determines its usefulness. The greater the time gap, the greater the uncertainty in census-based estimates. Other important factors affecting the usefulness of the census as a source of evidence of serious human rights crimes are (1) objectivity (of the statistical agency conducting the census), (2) coverage (the extent to which victimized populations are reached by the census), (3) bias (the extent to which the covered population correctly reports its attributes), and (4) availability (of the data files or databases for production of evidentiary tabulations or analysis).

In addition to being a source of aggregate statistical data, investigators may gain assistance in drawing samples of areas or persons for further study, or use the census to get information on individual victims. As we discuss below, the latter use can present ethical and legal problems. In some countries, the perpetrators have used

census data to help plan or to implement their crimes. In such cases, knowledge and understanding of census operations and outputs may be a significant contribution to the direct evidence of the crimes (Seltzer 1998, Seltzer and Anderson 2001).

Population censuses are the most common and established types of censuses, but countries also regularly conduct censuses on other topics, such as housing and businesses. Depending on circumstances, countries may conduct special censuses of population groups, such as refugees or minorities. The usual purpose of these (so-called) censuses of population groups is administrative, rather than statistical.

Another type of investigation, sometimes referred to as a “census of graves” or more properly grave-markers, may also provide useful data. Demographers have used information from tombstones (e.g., name, dates of birth and death, and information on family status) to develop useful estimates of mortality. However, such studies may have little relevance where bodies are disposed of through cremation or mass graves, individualized burial markers are not used or contain little identifying information, or the identifying information quickly weathers away.

Sample Surveys

Sample surveys are more varied in content, depth, auspices, and operations than population censuses. In many countries, the government, academic researchers, commercial organizations, and other groups conduct ongoing sample surveys of their population to gather data more frequently, more accurately, and in greater subject-matter depth than is possible in a regular population census. As in the case of the regular population census, the goal is to generate aggregate statistics, not information about named individuals. In addition to the ongoing survey programs, government, academic, and others conduct one-time, ad hoc surveys. The purposes of these surveys are usually determined by needs or interests of the moment and may be national or regional within a given country.

A number of technical issues arise in the collection and use of data based on samples. These include whether the sample is based on a probability sample design or a judgment sample, the specifications of the sample design, how well the sample design is followed in its execution, and the errors and uncertainties introduced because the study is based on a sample and not a full count. Often, the fact that data collection efforts are restricted to a sample of the population rather than the full population can mean that the interviewers are better trained and supervised, resulting in data of a higher quality. Assessing the various trade-offs involved can be a complex matter. An introduction to some of the issues and trade-offs involved is discussed in Section 10.2.2.3.

Administrative Registers

Administrative processes (e.g., birth and death registration, property regulation, and morbidity reporting) produce administrative data as a byproduct. The uses of, and methods of producing, such data are different from censuses and surveys.

Nonetheless, administrative data can be an important source of information about people in the aggregate and as individuals. In some regions, e.g., the Nordic countries, administrative registers are the primary source of most statistical information.

Some administrative records refer to people. People-oriented records include population registers, passport records, school records, driving license applications and issuances, and social security participant and benefit records. Other records refer to events, such as birth, marriage, death, divorce, and border crossing. Others refer to transactions, such as health, judicial, and criminal and prison records. Many of these registers may be of help in assessing the population at risk and the “normal” level of fertility, mortality, and migration. However, in periods of massive civil dislocation, often a feature of major human rights disasters, such administrative reporting systems often fail to function or function in a highly erratic fashion.

Also, since these systems may be in the hands of the perpetrators they may be run in a manner so as to disguise what actually happened. For example, despite the very detailed records kept by the Nazis during World War II, those gassed shortly after their arrival at the extermination camps were never individually registered at these camps; only those selected for work assignments were recorded as entering the camps. Similarly, the cause of death recorded on the death certificates issued for victims of human rights abuses may bear no relation to the actual cause of death. Finally, incriminating reports contained in administrative registers may be destroyed as part of a policy of obfuscation after their operational usefulness to the perpetrators is finished.

Despite these limitations, even fragmentary administrative records may prove highly useful to the ICC in conjunction with other evidence, quantitative and non-quantitative, in the investigation and prosecution of major human rights crimes. The ICC may also be able to make good use of private administrative records. Such records include data from private entities (such as funeral homes, retail stores, factories) or quasi-governmental organizations (such as postal services, communications carriers, and phone systems).

One category of administrative data stands out as directly and uniquely related to the events of concern: refugee, relief, and social services data. During and after the commission of major human rights violations, both governmental and non-governmental organizations may provide services and support to victims and refugees. In addition, governments often collect such data to control refugee populations. These sources of data play a key role in getting answers to the questions about the magnitude and nature of the crimes. For example, refugees may be created by forced displacement of virtually an entire well-defined subpopulation (as was done in Iraq to the Kurds in Anfal 1988). The missing components of the transported population (in that case, males of military age) provide evidence of systematic selection.

The ICC itself will generate administrative records from its own work and other sources, integrating data from a wide variety of sources. These records themselves are potentially complex, involving relational databases and the integration in networks—possibly international—of diverse databases to track complaints, alleged and verified violations and crimes, actions of perpetrators, events, military

equipment, and so forth. Ideally, the ICC could build a data warehouse,³ which could be queried to obtain ad hoc answers to ongoing questions concerning evidence. While human rights tribunals (or other NGOs) are unlikely to have the resources for development of a data warehouse on the scale of a commercial firm, experience shows that as such new technological approaches “trickle down,” the utilization costs decline in response to commercial development and use.

The diverse nature of potentially useful administrative reporting systems and different units of analysis (persons, crimes, unnamed cadaver discoveries) militates against generalization about either the systems or their data. At a minimum, however, we can observe that it is only the relevant data produced by such systems whose quality concerns us. For example, even if motor vehicle registrants habitually understate vehicle monetary values, the names and addresses of the registrants can still be a valid input to a cross-checking process for determining disappearances of individuals. The name and address data may have weight in a judicial setting because there is a legal basis for these variables of the registration records. Again, knowledge of the subject, the structure, and the characteristics of administrative records is essential. Moreover, such data sources may require considerable efforts at “data cleaning” to enable useful information to be extracted from them.⁴

This example shows the value of administrative systems databases that, by their very nature, are sources of data about named individuals, specific events, and transactions. They can give direct evidence of the status of persons prior to the violations of interest and generate high-quality lists to be used in more detailed investigations.

As in the case of censuses and surveys, objectivity, coverage, bias, and availability must be considered as potentially affecting uses of the data.

Medical and Anthropologic Forensic Studies

These studies produce a different type of quantitative data, which is based on autopsy findings, exhumations, etc. Such data often call for aggregation using quantitative methods (such as distributions of ages of victims in a mass grave) (Haglund et al. 2001, Steadman and Haglund 2005). Prior to the fieldwork, it may be appropriate to consider how sampling and other quantitative methods might be used to provide data that will be probative in the ICC environment.

³ Data warehousing is the process of acquiring, storing, and accessing enterprise data within an integrated database. It is an increasingly common technique in modern commercial information management. Many users can draw data from the warehouse through queries to analyze and explore the data.

⁴ Data cleaning involves the identification and correction of missing or inconsistent entries. Depending on the size and character of the data source, such data-cleaning operations may be very time-consuming and labor intensive, although some software is available to assist in aspects of this work. However, if not carefully implemented, data-cleaning operations, whether manual or computer-based, can produce very misleading results.

Case Studies

A case study is an intensive investigation of a specific population using quantitative and qualitative methods. In many countries, academic and governmental researchers conduct these studies on a continuing basis. ICC investigators can gain useful quantitative and qualitative information on populations and regions of interest to their work.

10.2.2.3 The Role of Sampling

Governments, industry, and academia make use of samples for one or more of the following reasons:

1. The cost or time needed to collect and compile data for the entire population (sometimes referred to as the “universe”) of interest exceeds the available resources.
2. Concentrating resources on gathering data from a smaller number of subjects or units improves the quality of the information gathered.
3. Measurement itself is destructive or difficult and thus cannot be carried out within the resource limitations (money, personnel, time, etc.) of the investigating body.

The ICC will find these reasons as cogent as do other users of data. Statistical sampling can make the Court’s investigative activities more timely, cost-effective, and precise. Table 10.2 provides some illustrative uses of sampling in the work of the ICC, in addition to the examples cited in Table 10.1 and elsewhere in the chapter.

A basic question to be resolved whenever sampling methods are used by the ICC (or on its behalf) as part of its investigative work or to gather evidence to be used in trial is whether to use a probability or non-probability sample or some combination of the two.⁵

In the short run, a non-probability (or “judgment”) sample is faster, requires fewer resources, and is less expensive. This approach may be appropriate in the

Table 10.2 Illustrative uses of sampling in the work of the ICC

<ul style="list-style-type: none"> ● Sampling to plan or carry out forensic and other investigative field studies, controlling for bias, and ensuring cost-effective use of resources and prompt action⁶ ● Selection of segments of sites for investigation (such as excavation, or survivors for interviews) ● Sampling to choose individual components (local administrative areas, dwelling units, or families) to generate macro- or micro-level data ● Follow-up procedures to get field lists of named persons previously known to have been alive ● Surveys of current dwelling occupants to improve demographic estimates of mortality, forced migration, or other forms of victimization.
--

⁵ In a probability sample, each element in the sample has known (or ascertainable) non-zero probability of selection. In a non-probability design, this is not the case.

⁶ We do not necessarily mean to imply probability sampling is used.

initial exploratory stages of the Court's investigations. However, cases and investigations before the ICC will usually involve complex processes about which there may be incomplete knowledge so that non-probability designs are best used with great care and for limited purposes. It is essential that non-probability samples be identified as such. Unfortunately, in the chaotic and uncertain conditions that usually follow the commission of gross human rights abuses, estimates based on non-probability or judgment samples can be the source of apparently plausible, but highly misleading findings. Moreover, the results of judgment samples are harder to defend than those of estimates based on probability samples. The defense may charge subjectivity as a challenge to the results of a judgment sample whereas a probability sample, which should be substantially free of selection and coverage bias, is usually less likely to be challenged on grounds of subjectivity. Finally, probability designs permit one to estimate the uncertainty introduced because information was gathered from a sample of units in the population rather than from the entire population.

These issues will be of greatest concern when the Court uses sample results to determine indictment or prosecution strategies or as evidence. In such cases, it can be an unwise economy to use a non-probabilistic sampling approach to hold costs down. However, investigators attempting a probability sample can encounter many serious practical problems, whose solutions may involve compromises with the soundness of the design of the sample or survey. The ICC staff will need to seek expert statistical advice to assist in assessing trade-offs that can affect the usefulness of the results obtained through sampling.

Probability sample designs range from simple to complex, in concept and execution. All other things being equal, the simplest satisfactory design is best in the context of the ICC. Not only are complex designs difficult to administer and costly, they would certainly be difficult to explain in a courtroom, which would offset any possible gain from the complexity. Many people have difficulty with statistical concepts; one common misconception being that a good sample is one that is some specified proportion of the sampled population.

One important problem that arises in the use of sampling to gather information from victim populations is securing a proper frame (i.e., a complete list of all elements potentially includable in the sample) required to select a sample. For example, surveys using a frame based on a listing of all households at the time of the survey would tend to underestimate mortality or migration because the survey could not reflect the experience of those households that no longer existed because all their members died or were forced to move away. The amount of bias from this source depends on the proportion of households falling into this category and their average size. When the populations of entire towns or villages are believed to have been wiped out, the problem is even greater. (Of course, even if sampled and interviewed, in a situation involving ongoing conflict, many respondents may be reluctant to provide full and accurate information out of fear of future retribution.)

During the course of fieldwork or the collection of data, situations will arise that make it difficult to adhere to the protocol of a sample design. For example, sampling methods can give useful guidance to forensic specialists excavating a mass grave that may call for making transects in a particular way. Practical considerations such

as the process of excavation, manner of burial, condition of bodies, embedded rocks and materials, and other subsurface conditions may make it impossible to follow the design protocol. Investigators and statisticians must balance simplicity, feasibility, costs, and scientific rigor. This is not as easy a compromise as in many industrial or research situations, and a significant loss of scientific rigor may render the results useless to the Court.

As when using a census, objectivity, coverage, bias, and availability can affect the ICC's use of sampling, including results based on sample surveys. However, in contrast to a census, it is possible to have multiple surveys of the same period, region, or other entity. With multiple surveys, it is possible to perform crosschecking to eliminate duplicate counts, and in some cases to estimate the number of total violations (CEH 1999, Annexo III). Moreover, unlike a national census, the ICC can readily initiate a sample survey at any time. Since the Court's concern may be in a limited region, this can be a reasonable approach even if only quite limited resources are available.

10.2.2.4 Methods of Quantitative Analysis

Descriptive Statistics

Descriptive statistics is the process of describing, presenting, and summarizing data to reveal or gain a better understanding about the processes that created the data. Exploratory Data Analysis (EDA) is a related set of techniques for understanding, analyzing, and presenting data, its structure and systematic patterns (Tukey 1977). Easily understood by non-professionals, EDA has a great deal to offer in a legal environment, despite an apparent lack of sophistication. Not the least of these advantages is demonstrated effectiveness, as shown by Hoaglin and Velleman (1995, p. 277):

Our examination shows that approaches commonly identified with Exploratory Data Analysis are substantially more effective [than a long list of advanced model-fitting methods] at revealing the underlying patterns in the data and at building parsimonious, *understandable* [our emphasis] models that fit the data well.

The summary measures of descriptive statistics and EDA play an important role in communicating aggregate characteristics. Statisticians and demographers use many measures of what is "typical" (averages such as means and medians), the dispersion of data (such as standard deviation and interquartile range), association, autocorrelation, and goodness of fit to a descriptive model.

Graphical representation is a principal method of descriptive statistics and EDA. The choice of graphical presentation calls for understanding the data, knowledge of graphical techniques, and recognition of the obligation to reveal without misleading. This is a task for a statistician, not an artist. A single adversarial challenge in a court could destroy the credibility of a presentation based on a misleading graphic.

Are the data credible? A statistician can use the tools of both descriptive and inferential statistics to find suspect data values. Are there outliers that suggest erroneous counting or measurement? Do the data have internal consistency? Of course,

it is best to avoid data quality problems from the start. As we have already stated, investigators can profit from the assistance of a statistician or demographer in planning data collection.

Inferential Statistics

Descriptive statistics makes no special assumptions about the nature of the data and the underlying processes. However, as the analysis of the data moves from description to exploration to confirmation, the statistician is led to make inferences about the process generating the data and the values of parameters of a sampled population (inferential statistics). The analyst must make assumptions about the processes and nature of the data collection. The strength—and sometimes the validity—of the results can be critically dependent on the fit between reality and the assumptions used in the analysis.

Inferential statistics has its theoretical basis in probability theory and as the quantitative analysis used in the court moves from description toward probabilistic evaluations (e.g., confidence intervals, hypothesis tests, likelihood ratios), it will encounter more difficulty. According to Meier et al. (1994, p. 24),

Tribe has argued that quantitative methods are easily misused: they often hide unstated assumptions, focus attention on those elements of a case most readily quantified, but not necessarily the most important. . . . these are all legitimate concerns, but they are equally applicable to any kind of expert testimony; experts often make unequivocal assertions later contradicted by other experts, state conclusions based on a large number of theoretical assumptions. . . . Just as with any form of technical and scientific evidence, the use of statistics calls for caution rather than neglect.

The leap from a statistical association to cause is frequently made. Can systematic statistical patterns of association be taken as probative in revealing causal relationships? In general, “Causation . . . is not usually settled by statistical arguments . . . though statistics has many inputs that can help” (Mosteller and Tukey 1977, p. 261). Although rule 93 of the rules of procedure and evidence for the ICTY states, “Evidence of a consistent pattern of conduct may be admissible in the interests of justice” (United Nations 1994a, p. 50), the equivalent rule for the ICC uses much more general language,

the Court may rule on the relevance or admissibility of any evidence, taking into account, inter alia, the probative value of the evidence and any prejudice that such evidence may cause to a fair trial or to a fair evaluation of the testimony of a witness...(ICC 2002, article 69, paragraph 4).

It is possible for statisticians to establish that systematic statistical patterns can meet courtroom standards for causation. In general, the criteria for causation include consistency, specificity, responsiveness, a plausible mechanism, elimination of possible common causes or confounding factors, strength of effect, and the impossibility of a reverse explanation for the apparent causal effect. For example, in a case involving theft from the City of New York, the judge allowed city attorneys to “demonstrate the pattern of theft as a legal matter through the introduction of

evidence” (DeGroot et al. 1994, pp. 79–106, 224). In this instance, the prosecution’s statisticians made a causal attribution by meeting a number of these criteria. The defense expert witness argued that reasonable alternatives to the prosecution causes could not be ruled out, but the court decided in favor of the prosecution (DeGroot et al. 1994, pp. 235, 240–246). In the courtroom, these issues are not likely to be resolved without the assistance of expert witnesses.

10.2.2.5 Processing and Compilation of Data

Investigators, attorneys, and their experts will want to retrieve and analyze data to meet their perceived legal needs. This process will be effective and efficient to the extent that the data are compiled to create an information management system to meet their needs. As Ball (1996, p. 1) says in his guide to implementing human rights information management systems,

The primary purpose of an information management system is to help a human rights organization understand and comprehend all of the information available to that organization. For example, if an investigator hears about a particular incident, she should be able to use the system to easily find other information about the incident that the organization has obtained in the past. . . . Finally, a good information management system should permit an organization to calculate statistics according to a variety of criteria. . . . The term “information management system” implies more than a computer database. Rather it suggests an integrated system through which an organization collects data, organizes it, puts it somewhere and then analyzes it.

We cannot reasonably expect investigators, attorneys, or even most scientists to know how to design information management systems and satisfactory methods of data storage. Commercial organizations employ highly paid specialists for this purpose. Unfortunately, human rights organizations have not always sought the assistance of personnel with current commercial-level skills. However, the ICC will have a larger staff and budget than these organizations and should have at least one permanent staff member expert in quantitative methods, with appropriate support personnel.

10.2.2.6 Ethical Standards

The use of scientific evidence in international criminal courts raises special ethical problems for the investigator, lawyer, and scientist. How does a statistical expert witness avoid being perceived as an advocate? When giving testimony, does the expert voluntarily reveal exculpatory analyses? Are there formalized standards for professional conduct of an expert witness that apply in this environment?

Under the Statute of Rome, the accused is guaranteed the right

To examine, or have examined, the witnesses against him or her and to obtain the attendance and examination of witnesses on his or her behalf under the same conditions as witnesses against him or her. The accused shall also be entitled to raise defences and to present other evidence admissible under this Statute (ICC 2002, article 67, paragraph 1(e)).

Thus, it possible that defense attorneys will seek statistical experts for assistance. If statisticians refuse to assist in the defense of an alleged perpetrator, should the ICC appoint one, and if appointed, should an individual serve?

The use of administrative or statistical data about individuals obtained under the promise of confidentiality (to the individual, the group, or the court) also raises ethical issues. Can prosecutors and investigators use such an individual population or administrative record for a victim (e.g., a population census record or an HIV/AIDS record, each collected under pledges of confidentiality), even if not using it would mean a failure to indict or convict?

The United Nations Statistical Commission has adopted The Fundamental Principles of Official Statistics (United Nations Economic and Social Council 1994, Chapter 5), which include the following:

... Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes.

A number of countries have specific laws that make it a criminal offense to divulge such data outside the statistical agency. The stress here is on “individual data,” as the legal and professional restrictions protect data pertaining to the individual. The rules of confidentiality do not usually apply to statistical aggregates. However, in the case of data collected by or for the ICC, there exist additional threats. Although aggregated data are by their nature stripped of explicitly identifying information, such as name, the identity of the data provider can sometimes be inferred by the defense or the perpetrator on the basis of the unusual characteristics of the data provider. For example, tabulations by educational attainment or occupation could permit the identification of the only person with advanced education or who was a medical doctor in some localities in a developing country.

Governments have used data collected with a promise of confidentiality to assist in major human rights abuses, including genocide and crimes against humanity. For example, the Nazi authorities in Germany, in the countries and areas it occupied, and in its client states used such data to commit crimes during World War II (Seltzer 1998). Does the use of such data in criminal activities justify the use of the same or other confidential official statistics in the investigation and prosecution of crimes? Moreover, the collection of data from victims and survivors during and after a human rights emergency even by well-meaning human rights workers can expose those interviewed to further abuses if the information provided falls into the hands of the perpetrators or their allies. Whether or not a government used confidential materials in committing crimes against humanity, do investigators and prosecutors have the right to use confidential materials in prosecution of those crimes? These are difficult questions that, in the long run, must be answered. What is the best form of representation to achieve standards for ethical considerations in this environment? Initial discussion of some of the issues involved may be found in Ball et al. (1997, p. 859), Seltzer and Anderson (2001), and Seltzer (2005).

A question also arises in connection with the ethical obligations relating to those who have already died. Clearly, in some circumstances, the identity of a dead person can pose risks to those still living (e.g., family members or others in

a targeted population subgroup). In other cases, the property rights or reputation of the deceased might be adversely affected. Any such risks would need to be taken into account in considering the continuing ethical obligations to protect confidentiality. Where no such risks exist, it would seem that a decision to permit identification would present minimal ethical concerns.

10.2.2.7 Education and Training

There is a clear need for at least some minimum training in quantitative methods for investigators and others working at the ICC. This need was articulated in a message sent to us some years ago, by an investigator working at one of the existing international criminal tribunals,

We [lawyers, investigators, and analysts preparing evidence] had a meeting a few days ago, and although we do a lot of disparate things we pretty much agreed that we're asked to provide analysis at the last moment and in a hodgepodge fashion. . . . We all seem to think there's something wrong in the way we are writing our reports, *but we don't have the training to do them better* [our italics]. For example, I just wrote a [report] . . . and I can't cross-reference anything, like how many people said X was shot and how many say that he was knifed, and out of those, how many said it was on Wednesday.

A similar need was identified by the U.S. National Academy of Sciences' Panel on Statistical Assessments as Evidence in the Courts when it recommended that

efforts be made to integrate instruction on basic statistical concepts into the law school curriculum and to provide instruction for practicing lawyers and judges on such matters. In law schools this can be done by allowing students to take for-credit courses outside the law school, by developing courses on statistics for lawyers, and by discussing statistical issues and concepts in specific courses. . . . For practicing attorneys and judges, such instruction may be provided by professional organizations, the Federal Judiciary Center, continuing education, and the like (Fienberg 1989, p. 182).

The Panel took a long-term view in calling for enhancing US legal education. However, how such training and education can be provided to those working at the ICC in the foreseeable future will be a challenge, given that it is largely absent from most personnel working in national criminal justice systems.

Some developments are encouraging. A number of American universities are now offering courses or special training programs for human rights workers that include at least minimum training in quantitative methods (e.g., Berkeley, Columbia, Johns Hopkins, Harvard, and Tufts). The content of these courses varies somewhat, but for an illustrative syllabus, see Spirer (1998). Some of those being trained are already working in international human rights organizations. The authors are hopeful that they will carry their knowledge of quantitative methods into their practice in human rights activities and, for a few, eventually to the ICC. It should be understood that such training programs will not make participants experts in statistics or demography. Indeed, an important element of such programs is to help non-technical staff to be aware of when and why expert advice and input is needed.

One of us (Spirer) has co-authored *Data Analysis for Monitoring Human Rights* (Spirer and Spirer 1993). The Ford Foundation, the American Association for the

Advancement of Science (AAAS), and the American Statistical Association (ASA) recognized the need when they sponsored the development of that handbook. Experience in instruction and consulting to NGOs quickly reveals the need for these skills. Unfortunately, while the need for consulting is recognized, often the staffs of the organizations do not seek education and training in this area.

Of course, some level of minimum training in quantitative methods is important for ICC lawyers and investigators, but the Court will also require access to those with advanced training and specialized experience in statistics and demography. An indication of the duties to be performed by such an expert and the qualifications required may be found in the job description used to hire a demographer at the ICTY:

Functions: Under the supervision of the Senior Research Officer, the incumbent will participate in current investigation processes, by undertaking demographic projects as set, liaise with both Prosecution and Investigation teams to establish prioritisation of projects for court purposes as well as authenticate data and assess data both aggregate and individual. He/she participates in briefing and planning conferences, prepares reports and briefs, participates when appropriate, in the presentation of prosecution briefs, testifies in formal proceedings of the Tribunal and performs administrative functions pertaining to the implementation of policies and procedures. Qualifications: University degree in demography or related field. Extensive experience in applied statistics required. Four to eight years of experience with applied demography or statistics, international experience would be desirable. Excellent computer skills (quoted in Brunborg 2001).

Such experts, whether serving on the ICC staff, working on commissioned studies for the Court, engaged as consultants, or serving as expert witnesses, will also require an orientation into the special requirements of the legal setting in which they will work. Indeed, all scientists regardless of their discipline need training to be effective in serving legal purposes. Any litigator can give anecdotes of difficulties in their uses of expert witnesses, often resulting from a lack of understanding of the role of science in the court. Here is how one scientist described the issue:

[Scientists] like to think of science as a tool for understanding the universe. In a courtroom, it is harnessed to a need to find winners and losers. It is all black and white, there are no shades of gray. . . (Foote 1996).

10.2.2.8 The Costs of Statistical and Demographic Work

The range of possible costs of specific statistical and demographic activities for the ICC is wide, from modest to quite high. How best to arrange for the expert services needed will in part depend on the size and nature of the case load of the ICC. In the absence of specific information on these and related matters, we can only speak in general terms.

The costs of analysis can be expected to fall in the modest range most of the time. Middle or senior level statistical or demographic professionals can do most of the analytical projects for a specific case in a period of several months. However, some analytical tasks can be demanding, time-consuming, and require specialized knowledge. For example, if the prosecution must make a case about crimes which involved large numbers of people, the process of making the needed verifications

can call for the services of documentation specialists, statisticians, and database designers, as well as persons familiar with particular languages and cultures.

Field work of any kind makes high demands for resources. In the example above, fieldwork may have to be coupled with the ongoing analysis. The investigations needed to collect credible data call for a variety of experts with rare skills (language, database, records organization) and a talent for tracking down information that is deliberately withheld or difficult to obtain. For example, records can have been deliberately or inadvertently destroyed or simply deteriorated in an adverse environment. Cadaver remains can have been chemically destroyed or moved. In Rwanda, many skeletal remains were moved and then entombed in concrete for memorials. There are many such known cases and investigators may need many months to achieve modest incremental results.

If the goal is to make reliable and credible national estimates of the magnitude and scope of violations and victims with data gathered from a newly commissioned probability sample, costs can range from several hundred thousand to millions of US dollars. On the other hand, if use can be made of demographic data gathering carried out by other agencies (e.g., a national population census or a national household survey), costs of data acquisition to the ICC can be much more modest. Factors affecting costs of field work carried out by the ICC, or on its behalf, include, inter alia, the size and distribution of the surveyed population, the size of the country, the country's communication infrastructure, the availability of existing records, the nature and desired precision of estimates, and the need for specialized support and staff to meet the needs of political credibility.

10.2.2.9 Standards for Assessing Statistical Evidence

The US Supreme Court, in commenting on criteria for evaluating scientific evidence in US courts, observed:

Science is a multifaceted discipline, the potential for error or abuse quite variegated, and its application in litigation quite diverse. Broad standards . . . will inevitably require substantial amplification with more specific and contextual principles (Foote 1996).

We believe that this statement is valid for the scientific sub-disciplines of statistics and demography, and that the potential applications are indeed diverse and varied. The ICC will need to develop its own standards for the evaluation of statistical evidence through its ongoing operations.

10.2.2.10 The Respective Roles of Lawyers, Investigators, Scientists, and the Court

To some extent, the legal, law enforcement, and scientific viewpoints on the use of quantitative methods and data derives from the differences in purpose. The legal and law enforcement goals are to find out who (if anyone) did what (if anything) to whom (if anyone) how many times. In the final analysis, a full telling of the truth is

secondary to its use in arrest, indictment, prosecution, and conviction. On the other hand, science “searches for truth and seeks to increase knowledge by formulating and testing theories” (Fienberg 1989, p. 140).

In the courtroom environment, the several parties (court, prosecution, and defense) make decisions on individual cases. Quantitative analysis almost invariably is only one of a number of inputs; it rarely stands by itself. When there is clear evidence of motivation, intent, and criminal action such as from testamentary evidence, quantitative analysis may play no role. And in some cases, a scientist may be troubled when one dramatic anecdotal report has the dominant impact on proof of guilt.

The role of the statistician or demographer will be to develop, assemble, and perhaps present data on the number of victims, their characteristics, and any relevant information on the pattern of mortality or migration. The relevancy of such scientific findings will be an issue for the legal actors, not the expert scientific witness. A scientist may find a result convincing beyond a shadow of a reasonable doubt, but it is the prosecutor or defense attorney who decides on how to use the result in court, if it is used at all. The Court itself will have the final word. For example, the rules of evidence used in the ICTY provide that the Tribunal can “exclude evidence if its probative value is substantially outweighed by the need to ensure a fair trial” (United Nations, 1994a). A scientist may find this hard to accept.

10.3 Experience in the Use of Demographic and Statistical Data in International Criminal Tribunals

Modern international experience in prosecuting perpetrators of serious human rights violations is limited to (a) the series of trials of those charged with war crimes and crimes against humanity that took place at the end of World War II, (b) the ICTR, and (c) the ICTY. We examine the use of demographic and statistical data and methods in each of these situations. In the case of prosecutions at the end of World War II, we confine our examination to the Trial of the Major War Criminals carried out in Nuremberg in 1945 and 1946. For the purposes of presentation, we discuss the ICTR before the ICTY. However, we note that the ICTY was established in May 1993 about 18 months prior to the creation of the ICTR in November 1994. Both the ICTR and the ICTY are under instructions from the United Nations Security Council (Resolution 1503 of 2003) to complete their work by 2010.

In considering the experience of the ICTR and ICTY, in relation to the ICC, it is useful to keep in mind one important distinction between legal philosophy governing the ad hoc tribunals on the one hand and the ICC on the other. The ICTR and the ICTY are based largely on a common law tradition as practiced in the United States and the United Kingdom, while the ICC is largely based on a civil law tradition. The former system is essentially adversarial in nature, while under the latter system, the prosecutor’s charge is to present all relevant evidence to the Court, whether it supports conviction or is exculpatory. Although this and other differences in the two

approaches are not expected to have as much impact on the decisions of the ICC in practice, since judges under the ICC and the ad hoc tribunals were appointed from jurists who had worked under both traditions, it may influence the work of the prosecutor's office.

10.3.1 Nuremberg

In his opening statement at the Nuremberg trials, Robert Jackson, chief US prosecutor, in addressing "Crimes against the Jews" under count one of the indictment, not only referred to figures pertaining to the numbers who had disappeared, but also introduced some analytical discussion of them:

I advert to [these figures] only to show their magnitude as evidence of a purpose and a knowledge common to all defendants, of an official plan rather than of a capricious policy of some individual commander, and to show a continuity of Jewish persecution from the rise of the Nazi conspiracy to its collapse as forbids us to believe that any person could be identified with any part of Nazi action without approving this most conspicuous item in their program (as quoted in Seltzer, 1998).

Seltzer also noted that as the Trial progressed additional quantitative data on the overall losses became available to the Prosecution and that these new estimates were then referred to in the Prosecution and witness statements to the Tribunal. It appears that no demographers or statisticians were called as expert witnesses at Nuremberg. However, Seltzer (1998, p. 533) considered that it seemed clear that "some sort of expert demographic advice" was available to the Chief Prosecutor since he used the language of technical demography to explain that his initial estimate of over 4.5 million losses was "beyond those accounted for by normal mortality and emigration." Finally, the Tribunal in its judgments repeatedly made reference to figures on the estimated numbers of persons killed.

10.3.2 International Criminal Tribunal for Rwanda (ICTR)

One of us (Seltzer) served as a consultant to the Office of the Prosecutor of the ICTR in 1996 and offered a set of recommendations on the use of data and quantitative methods in the work of the Tribunal. These recommendations were essentially ignored and the ICTR did not attempt to formally assemble quantitative evidence or commission demographic studies to complement the other evidence before the Tribunal. However, individual expert witnesses did consult with demographers to obtain knowledgeable information about the soundness of various overall estimates of genocide mortality and to receive guidance on the availability and interpretation of data from the 1991 Rwandan Census (DesForges 1999).

The additional role that statistical and demographic data and analysis could have played in the work of the ICTR can be demonstrated by several of the Tribunal's rulings in individual cases and some of the studies that have been published by

demographers working independently from the ICTR. For example, in one case, Akayesu (ICTR-96-4), the Trial Chamber observed that “intent,” which is a critical element in the crime of genocide, and “which is difficult, even impossible to determine,” could be inferred from several factors, including, among other things, “the scale of atrocities committed” and “the fact of deliberately and systematically targeting victims on account of their membership of a particular group, while excluding the members of other groups” (Judgement, paragraphs 523–534, cited in Human Rights Watch 2004, p. 14). Similarly, in Kayishema and Ruzindana (ICTR-95-1), the Trial Chamber found that “intent can be inferred from either words or deeds and may be demonstrated by a pattern of purposeful action.” Among the seven relevant indicators specified, two were directly based on quantitative data: “the number of group members affected” and “the relative proportionate scale of the actual or attempted destruction of a group” (Judgement, paragraphs 93 and 527, cited in Human Rights Watch, 2004, p. 14). Independent demographic and statistical studies may be illustrated by the work of Verwimp (2003, 2004) and Verpoorten (2005).

10.3.3 International Criminal Tribunal for the Former Yugoslavia (ICTY)

One of us (Spirer) also served as a consultant to the Office of the Prosecutor of the ICTY and offered a parallel set of recommendations on the use of data and quantitative methods in the work of that Tribunal. These recommendations were slow to be acted upon, and in April 1997 the authors convened an interdisciplinary conference at Columbia University in New York to bring together investigators, lawyers, statisticians, and demographers to promote a clearer understanding of how quantitative data could serve the needs of international criminal tribunals (Spirer, 1997). In due course, however, a number of recommendations made to develop and use demographic and statistical data were implemented by the ICTY, although not necessarily as a result of the April meeting. Two important initiatives may be cited.

First, the ICTY engaged the services of a senior statistician-demographer, Helge Brunborg, on loan from Statistics Norway. Building on Brunborg’s work, the ICTY created a Demographic Unit in the Office of the Prosecutor, headed by Ewa Tabeau, an experienced statistician-demographer. After returning to Norway and working in cooperation with this Unit, Brunborg was able to assist the ICTY in a number of ways, including helping the Tribunal’s lawyers and investigators understand the possibilities and limitations of quantitative data and advising on the application of sound approaches to additional data gathering and to the establishment of an extended quantitative capability in the ICTY (Brunborg 2001). His paper also describes some of the estimates of losses in Bosnia and Herzegovina during 1992–1995. However, his most notable contributions to the work of the ICTY and international human rights law was the careful work he and his assistants carried out in developing estimates of those Bosnian Muslim males killed in Srebrenica (Brunborg and Urdal, 2000, Brunborg et al. 2003).

The clearest way of demonstrating Brunborg's impact is to simply quote from the judgments in the ICTY Trial Chamber and the Appeals Chamber in the *Prosecutor v. Radislav Krstic*, IT-98-33 "Srebrenica," issued in August 2001 and April 2004, respectively, and available from the ICTY website (www.un.org/icty/). The Trial Chamber in setting out its findings of fact cites Brunborg's work and testimony in several places, including the observation that

the sex distribution of the persons listed as missing from Srebrenica, on the International Committee of the Red Cross (ICRC) list (cross referenced with other sources), correlates with the sex distribution of the bodies exhumed from the graves. Professor Helge Brunborg, a Prosecution demographics expert, testified that the overwhelming majority of people registered as missing from Srebrenica are men ... Similarly, there is a correlation between the age distribution of persons listed as missing and the bodies exhumed from the Srebrenica graves: 26.4 percent of persons listed as missing were between 13–24 years and 17.5 percent of bodies exhumed fell within this age group; 73.6 percent of persons listed as missing were over 25 years of age and 82.8 percent of bodies exhumed fell within this age group (paragraph 74).

Although noting that "it is impossible to determine with precision the number of Bosnian Muslim men killed by Bosnian Serb forces following the take-over of Srebrenica in July 1995," the Trial Chamber Judgment first referred to the limited forensic evidence available and then continued:

The number of people still listed as missing from Srebrenica in July 1995 provides further guidance as to the likely number of men executed. Professor Brunborg testified that, conservatively estimated, a minimum of 7,475 persons from Srebrenica are still listed as missing, based on the cross-referencing of ICRC lists and other sources and that it is likely that the vast majority of these missing people are deceased (paragraph 81).

It is clear that Brunborg's studies and testimony were central to the Tribunal's conclusion that "in July 1995, following the take-over of Srebrenica, Bosnian Serb forces executed several thousand Bosnian Muslim men. The total number is likely to be within the range of 7,000–8,000 men" (paragraph 84) and its ultimate judgment was that defendant Krstic was guilty of the crime of genocide (ICTY, Press release No. 609).

Although the Appeals Chamber ultimately decided that Krstic was not guilty of genocide as such, but of the somewhat lesser crime of aiding and abetting genocide, it unanimously found that "genocide was committed in Srebrenica in 1995" (ICTY, Press release No. 839). The Appeals Chamber, in considering whether the killings in Srebrenica constituted genocide, observed:

The intent requirement of genocide under ... the Statute is therefore satisfied where evidence shows that the alleged perpetrator intended to destroy at least a substantial part of the protected group. The determination of when the targeted part is substantial enough to meet this requirement may involve a number of considerations. The numeric size of the targeted part of the group is the necessary and important starting point, though not in all cases the ending point of the inquiry. The number of individuals targeted should be evaluated not only in absolute terms, but also in relation to the overall size of the entire group. In addition to the numeric size of the targeted portion, its prominence within the group can be a useful consideration. If a specific part of the group is emblematic of the overall group, or is essential to its survival, that may support a finding that the part qualifies as substantial within the meaning of [the Statute] ...(paragraph 12).

This judgment is an important one for international criminal law in several respects. First, it establishes that even the killing of 7,000–8,000 persons in the appropriate circumstances can constitute genocide. Furthermore, its finding that “the numeric size of the targeted part of the group is the necessary and important starting point, though not in all cases the ending point of the inquiry” makes clear that quantitative data are important for work of the Tribunal and its further reference of the need to take into account the relative size of the targeted group and “its prominence within the group” also make clear that in addressing quantitative data there is a need to involve skilled analysts. Finally, the Appeals Chamber accepted without question the Trial Chambers quantitative findings, based largely on Brunborg’s work.

The ICTY made use of another expert witness, Patrick Ball, who testified for the Prosecution in the Kosovo phase of *Prosecutor v. Slobodan Milošević* (IT-02-54) on March 13 and 14, 2002 and May 2, 2003. His testimony was largely a presentation of a report, *Killings and Refugee Flow in Kosovo March–June 1999* (Ball et al. 2002a,b), prepared for the Tribunal. The report made use of advanced statistical techniques (e.g., multiple system estimation using several imperfect data sources) to prepare estimates of overall mortality and forced migration to support the charge of crimes against humanity against Milošević for actions in Kosovo. Even more telling, Ball and his colleagues examined the observed spatial and temporal patterns of estimated mortality and migration in light of the various theories offered by the defense (i.e., NATO bombing and military actions of the Kosovo Liberation Army) to explain the mortality and migration that took place.⁷ Since no final judgment can be rendered, we cannot be clear about the impact of this statistical and demographic work on the Tribunal’s findings and conclusions in the Milošević case.

However, the overall impact of the Demographic Unit within the Office of the Prosecutor of the ICTY is already clear. In 2005, according to Ewa Tabeau (2005), the Unit consisted of its director, a database specialist, and a clerk. Since 2000, the group, assisted by such experts as Helge Brunborg, Patrick Ball, and others, has produced over 15 expert reports, excluding small analytical studies, many of which have been introduced as evidence in individual cases (see the listing in Tabeau et al. 2003, Annex D of reports issued as of March 31, 2003). A selection of these expert reports is also individually cited in the references for this chapter (see, Brunborg 2002 and 2003, Tabeau and Zoltkowski 2002a and b, and Tabeau et al. 2002). In addition, demographers and statisticians have testified as Prosecution expert witnesses in eight cases up to November 2005, either to advance the Prosecution’s case or to rebut expert testimony offered by the defense.

In addition to the Krstic case described above, several other judgments issued by the Tribunal have also made explicit references to the expert reports prepared by the Demographic Unit and its statistical and demographic consultants (see, e.g., ICTY judgments in Vidoje Blagojevic and Dragan Jokic, IT-02-60, paragraph 291; Stanislav Galic, IT-98-29-T, paragraphs 578–581; Biljana Plavsic, IT-00-39&40/1,

⁷ Brunborg also testified as an expert witness for the Prosecution in the Croatia and Bosnia phase of *Prosecutor v. Slobodan Milošević* (IT-02-54), “Kosovo, Croatia and Bosnia,” Transcript, 2/11/2004, pp. 31936–31954.

paragraphs 36–39; Blagoje Simic, Miroslav Tadic, and Simo Zaric, IT-95-9-T, paragraphs 30–35; Milomir Stakic, IT-97-24-T, paragraphs 553 and 646; and Mitar Vasiljevic, IT-98-32-T, paragraph 39). It may be noted that, with the sole exception of the Stakic case, the Tribunal not only referred to the evidence developed by the Unit, but considered it to be part of the factual basis of its judgment.

10.4 Discussion and Conclusions

Clearly, any measures that increase the cost-effectiveness of investigative and evidentiary activities will be important to the ICC. In the preceding sections we have identified and discussed the issues that we feel are relevant to the effective and efficient use of data and quantitative analysis at the ICC. In this summary, we put forward questions raised by our discussions in the hope that concerned lawyers, statisticians, and demographers will join in public discussion which will lead to contributions and recommendations that will help to support the work of these tribunals.

Based on national experience and experience with the ICTR and ICTY we have reached two firm conclusions. First, the ICC will need access to an international community of experts, including those with expertise in statistics and demography and related training materials and services. Secondly, there is a continuing need to raise awareness of statistical and demographic data and methods and their legal use among investigators, lawyers, prosecutors, defense attorneys, and judges. We hope that the experience recounted in this chapter will contribute to a wider recognition by human rights lawyers and others associated with the ICC of the value of quantitative data and analysis to the work of the ICC.

By definition, the ICC is an endeavor involving victims, perpetrators, defendants, attorneys, analysts, investigators, and judges from the world at large, and we can expect rules of evidence that reflect many different systems of justice. The industrialized nation-states tend to dominate the human rights field and the applications of quantitative methods and data to solving problems. Their size, wealth, infrastructure, and educational system support this dominance. Added to the problem of crossing the professional boundary between the legal and quantitative disciplines, we have to deal with a multiplicity of international styles, laws, and languages.

If the contributions of data and quantitative analysis are to have value, then it will be the investigators, lawyers, and judges who initiate most of the joint working relationships. Statisticians or demographers cannot on their own initiative place themselves on the stand as expert witnesses, working for the ICC or the prosecution. Of course, as we have discussed above, if the legal staff is not aware of the value of collaboration, then there will be few or no joint working relationships. In our experience, where there is such awareness, there is little trouble at the working level integrating the several disciplines and cultures. Success of education and training programs such as we describe will be realized in more effective and efficient joint efforts.

We end with this final thought. All of us who are involved in the work of an international criminal court—statisticians, demographers, investigators, prosecutors,

judges—should have but one goal: to bring justice and closure for the families of victims and the survivors of genocide and other gross crimes against humanity. Finding the best way to meet this goal must be our joint, our cooperative, and our imperative mission.

Acknowledgment The present chapter is a substantially revised and updated version of a paper, “Obtaining Evidence for International Criminal Tribunals Using Data and Quantitative Analysis,” that the authors prepared in connection with a one-day conference they organized in April 1997 at Columbia University on that topic. The authors gratefully acknowledge the assistance of Tom Jabine, statistical consultant, Louise Spierer, independent scholar, and John Ralston, former chief of Investigations, Office of the Prosecutor, International Criminal Tribunal for the Former Yugoslavia (ICTY) in preparing the original paper. They are also grateful to Helge Brunborg of Statistics Norway, Morton Bergsmo formerly of the ICC, and Ewa Tabeau of the ICTY for comments made in their personal capacity on an earlier version of this chapter as well as the helpful comments of two anonymous reviewers.

References

- Ball P (1996) Who did what to whom? Planning and implementing a large scale human rights data project. American Association for the Advancement of Science (AAAS), Washington
- Ball P (2000) Policy or panic? The flight of ethnic Albanians from Kosovo, March–May 1999. AAAS, Washington
- Ball P, Betts W, Scheuren F, Dudukovich J, Asher J (2002a) Killings and refugee flow in Kosovo March–June 1999. A report to the International Criminal Tribunal for the Former Yugoslavia. AAAS and American Bar Association Central and East European Law Initiative, Washington
- Ball P, Betts W, Scheuren F, Dudukovich J, Asher J (2002b) Addendum to expert report of Patrick Ball. AAAS, Washington (http://hague.bard.edu/reports/hr_ball-6mar2002.pdf, accessed August 16, 2006)
- Ball P, Girouard M, Chapman A (1997) Information technology, information management, and human rights: a response to Metzler. *Human Rights Quarterly* 19:836–859
- Ball P, Kobrak P, Spierer H (1999) State violence in Guatemala, 1960–1996: a quantitative reflection. AAAS and Centro Internacional por Investigaciones en Derechos Humanos, Washington
- Besson J-L, dir (1992) *La Cité des chiffres: Ou l’illusion des statistiques*. Editions Autrement, Paris
- Brunborg H (2001) Contribution of statistical analysis to the investigations of the International Criminal Tribunals. *Statistical Journal of the United Nations Economic Commission for Europe* 18:227–238.
- Brunborg H (2002) Report on the size and ethnic composition of the population of Kosovo: case of Slobodan Milošević (IT-02-54). ICTY, The Hague (http://hague.bard.edu/reports/Helge_Brunborg.pdf, accessed August 16, 2006)
- Brunborg H (2003) Needs for demographic and statistical expertise at the International Criminal Court: contribution to an expert consultation process on general issues relevant to the ICC Office of the Prosecutor. ICTY, The Hague (www.icc-cpi.int/library/organs/otp/brunborg.pdf, accessed August 16, 2006)
- Brunborg H, Lyngstad TH, Urdal H (2003) Accounting for genocide: how many were killed in Srebrenica? *European Journal of Population/Revue européenne de Démographie* 19: 229–248
- Brunborg H, Urdal H (2000) Report on the number of missing and dead from Srebrenica: expert report prepared for the Krstić case (IT-98-33). International Criminal Tribunal for the Former Yugoslavia (ICTY), The Hague
- Comisión para el Esclarecimiento Histórico (CEH) (1999) Metodología intermuestra: Documento elaborado por la Asociación para el Avance de la Ciencia. In CEH, *Memoria del Silencio*. CEH, Guatemala

- Coalition for the International Criminal Court (CICC) (1996–1998, 2002–2005) International Criminal Court Monitor. World Federalist Association and CICC, Washington (www.iccnw.org/index.php?mod=monitor, accessed August 16, 2006)
- DaPonte BO (1993) A case study in estimating casualties from war and its aftermath: the 1991 Persian Gulf War. *Physicians for Social Responsibility Quarterly* 3:57–66
- DeGroot MH, Fienberg SE, Kadane JB (eds) (1994) *Statistics and the law*. John Wiley and Sons, New York
- DesForges AL (1999) *Leave none to tell the story: genocide in Rwanda*. Human Rights Watch, New York
- Demographic Unit of the ICTY (2003) *The Bosnian war in numbers: overall war-related mortality, internally displaced persons and refugees from Bosnia and Herzegovina*. Presentation at special session on demographic consequences of the 1990s conflict in Bosnia and Herzegovina. European Population Conference, Warsaw (August 27, 2003)
- Federal Judiciary Center (1994) *Reference manual on scientific evidence: for use with federal practice and procedure*. West Publishing Company, St Paul
- Fienberg SE (1972) The multiple recapture census for closed populations and incomplete 2k contingency tables. *Biometrika* 59:591–603
- Fienberg SE (1992) Bibliography on capture recapture modeling with application to census undercount adjustment. *Survey Methodology* 18:143–154
- Fienberg SE, ed (1989) *The evolving role of statistical assessments as evidence in the courts*. Springer-Verlag, New York
- Fienberg SE, Johnson MS, Junker BW (1999) Classical multilevel and Bayesian approaches to population size estimation using multiple lists. *Journal of the Royal Statistical Society, Series A* 162:383–405
- Foote SB (1996) Book Review of Green MD (1996) *Benedictin and birth defects*. *Science*, 273:196
- Frelick B (1988) *Refugees: contemporary witnesses to genocide*. Presentation at genocide watch conference. Institute for the Study of Genocide at John Jay College of Criminal Justice, New York
- Gastwirth JL (1997) Book review, *Reference manual on scientific evidence*. *The American Statistician* 51:297–299
- Haglund W, Connor M, Scott D (2001) The archaeology of contemporary mass graves. *Journal of Historical Archaeology* 35:57–69
- Hoaglin DC, Velleman PF (1995) A critical look at some analyses of major league baseball salaries. *The American Statistician* 49:277–285
- Human Rights Watch (2004) *Genocide, war crimes, and crimes against humanity: topical digests of the case law of the International Criminal Tribunal for Rwanda and the International Criminal Tribunal for the Former Yugoslavia*. Human Rights Watch, New York
- International Criminal Court (ICC) (2002) *Rome Statute of the International Criminal Court*. ICC, The Hague (www.icc-cpi.int/library/about/officialjournal/Rome_Statute_120704-EN.pdf, accessed August 16, 2006)
- ICC(2005) Decision of pre-trial Chamber I on the prosecutor's request under article 56 of the Rome statute: press release on April 26, 2005. ICC, The Hague (www.icc-cpi.int/press/press-releases/105.html, accessed August 16, 2006)
- ICC, Assembly of States Parties (2002) *Rules of procedure and evidence adopted by the assembly of states parties, first session, New York, September 3–10, 2002*, Official records ICC-ASP/1/3. ICC, The Hague (www.icc-cpi.int/library/about/officialjournal/Rules_of_Proc_and_Evid_070704-EN.pdf, accessed August 15, 2007)
- International Criminal Tribunal for the Former Yugoslavia (ICTY) Index to trial and appellate chamber judgments available at www.un.org/icty/cases-e/index-e.htm and index to press release archive available at www.un.org/icty/latest-e/index.htm, both accessed August 15, 2007.
- International Criminal Tribunal for Rwanda (ICTR) Index to trial and appellate chamber judgments available at <http://trim.unict.org/index1.htm>, both accessed August 15, 2007.
- Louis-Jacques L, (ed) (2003) *International Criminal Court: resources in print and electronic format*, updated August 12, 2003 (<http://lib.uchicago.edu/~llou/icc.html>, accessed June 23, 2005)

- Marks ES, Seltzer W, Krotki KJ (1974) Population growth estimation: A handbook of vital statistics measurement. The Population Council, New York
- Meier P, Sacks J, Zabell S (1994) What happened in Hazelwood. In: DeGroot MH, Fienberg SE and Kadane JB (eds) Statistics and the law. John Wiley and Sons, New York
- Mosteller F, Tukey JW (1977) Data analysis and regression. Addison-Wesley, Reading
- Roberts S, Williams J (1995) After the guns fall silent: The enduring legacy of landmines. Vietnam Veterans of America Foundation, Washington
- Robertson B, Vigneaux GA (1995) Interpreting evidence: evaluating forensic science in the courtroom. John Wiley and Sons, New York
- Samuelson D, Spierer HF (1992) Use of incomplete and distorted data in inference about human rights violations. In: Jabine TB and Claude RP (eds) Human rights and statistics: getting the record straight. University of Pennsylvania Press, Philadelphia
- Seltzer W (1998) Population statistics, the holocaust, and the Nuremberg trials. *Population and Development Review* 24:511–552
- Seltzer W (2005) Statistics and counterterrorism: the role of law, policy and ethics. Presented at the workshop on statistics and counterterrorism, National Institute of Statistical Science and the American Statistical Association, New York, November 20, 2004, 2005 Proceedings of the American Statistical Association, Section on Risk Analysis [CD-ROM], Alexandria, VA: American Statistical Association, pp. 4052–4059.
- Seltzer W, Anderson M (2001) The dark side of numbers: the role of population data systems in human rights abuses. *Social Research* 68:481–513
- Spieler L (1997) 12 April 1997 meeting issues discussed (using data and quantitative analysis to obtain evidence for international criminal tribunals) (December 20, 1997)
- Spieler H (1998) Syllabus for U8165, workshop in human rights field research, data analysis, and reporting., updated September 1998
- Spieler H, Spieler L (1993) Data analysis for monitoring human rights. AAAS and HURIDOCS, Washington
- Steadman DW, Haglund WD (2005) The scope of anthropological contributions to human rights investigations. *Journal of Forensic Sciences* 50:1–8
- Tabeau E (2002) Basic demographic characteristics and socio-economic status of missing and killed persons from the municipality of Prijedor, 30.04-30.09.1992: expert report prepared for the Stakic case (IT-97-24). ICTY, The Hague
- Tabeau E (2005) Personal communication between W Seltzer and E Tableu, November 17, 2005
- Tabeau E, Bijak J (2005) War-related deaths in the 1992–1995 armed conflicts in Bosnia and Herzegovina: a critique of previous estimates and recent results. *European Journal of Population* 21:187–215
- Tabeau E, Zoltkowski M (2002a) Demographic consequences of the conflict in the municipality of Vlasenica, May–September 1992: expert report prepared for the Nikolic case (IT-94-2-PT). ICTY, The Hague
- Tabeau E, Zoltkowski M (2002b) Ethnic composition and displaced persons and refugees in 37 municipalities of Bosnia and Herzegovina, 1991 and 1997: expert report prepared for the Krajisnik-Plavsic case (IT-00-39&40). ICTY, The Hague
- Tabeau E, Zoltkowski M, Bijak J (2002) Population losses in the Siege of Sarajevo, September 10, 1992 to August 10, 1994: expert report prepared for the Galic case (IT-98-29- I). ICTY, The Hague
- Tabeau E, Zoltkowski M, Bijak J, Hetland A (2003) Ethnic composition in and internally displaced persons and refugees from 47 municipalities of Bosnia and Herzegovina, 1991 and 1997: expert report for the Case of Slobodan Milošević (IT-02-54). ICTY, The Hague (http://hague.bard.edu/reports/Ewa_Tableu.pdf, accessed August 16, 2006)
- Tufte E (1997) Visual explanations: images and quantities, evidence and narrative. Graphics Press, Cheshire
- Tukey JW (1977) Exploratory data analysis. Addison-Wesley Publishing Co, Reading
- United Nations Economic and Social Council (1994) Report of the special session of the statistical commission (New York, April 11–15, 1994), E/1994/29. United Nations, New York

- United Nations (1983a) Indirect techniques for demographic estimation – manual X. United Nations, New York
- United Nations (1983b) Mortpak-lite: the United Nations software package for mortality measurement. Population studies No. 104. United Nations, New York
- United Nations (1990) Mortpak and Mortpak Lite upgrades. Population studies No. 117. United Nations, New York
- United Nations (1992) General Assembly, report of the International Law Commission on the work of its forty-fourth session (A/RES/47/33), November 25, 1992. United Nations, New York
- United Nations (1993) General Assembly, report of the International Law Commission on the work of its forty-fifth session (A/RES/48/31), December 9, 1993. United Nations, New York
- United Nations (1994a) Rules of procedure and evidence, International Tribunal for the Prosecution of Persons Responsible for Serious Violations of International Humanitarian Law Committed in the Territory of the Former Yugoslavia since 1991 Second Session (IT/32), March 14, 1994. United Nations, New York
- United Nations (1994b) General Assembly, establishment of an international criminal court (A/RES/49/53), December 9, 1994. United Nations, New York
- United Nations (1995) General Assembly, establishment of an international criminal court (A/RES/50/46), December 11, 1995. United Nations, New York
- United Nations (1996) General Assembly, establishment of an international criminal court (A/RES/51/207), December 17, 1996. United Nations, New York
- United Nations (1997) General Assembly, establishment of an international criminal court (A/RES/52/160), December 15, 1997. United Nations, New York
- United Nations (2004) Death and survival during the 1994 genocide in Rwanda. Population Studies 58:233–245
- Verpoorten M (2005) The death toll of the Rwandan genocide: a detailed analysis for Gikongoro Province. Population (English edition) 60: 331–368
- Verwimp P (2003) Testing the double-genocide thesis for central and southern Rwanda. Journal of Conflict Resolution 47:423–442
- Verwimp p(2004) Death and Survival during the 1994 genocide in Rwanda. Population studies 58:233–245

Chapter 11

New Issues in Human Rights Statistics

David L. Banks and Yasmin H. Said

Abstract The traditional Western conception of human rights has focused upon political and civil liberties. Recent trends are bringing social and economic rights to the fore. This chapter considers the kinds of contributions that statisticians might make in this area and how issues of equity might be framed from a mathematical perspective. The emphasis is upon identifying problems and proposing methods, rather than analysis of specific applications.

11.1 Introduction

When the statistics of human rights began, the cold war was hot. Most attention focused upon civil and political rights—the right to vote, the right to freedom of religion, and the right to citizenship. But there has long been another thread, holding that people have a right to food, education, and medicine. This chapter focuses upon the statistical aspects of these social and economic rights.

Obviously, statisticians can play and have played a major role in tracking trends in the standard of living over time. We have participated in model building for economic input–output models that speak to the mechanisms of improvement. We have helped to invent quality of life measures that are pertinent to this area, and we have documented the impact of war on civilians. And there are new challenges that are arising in this arena—the 2005 G8 summit is trying to rescue Africa from poverty, the eight Millennium Development Goals (see www.un.org/millenniumgoals/) require statistical documentation to support funding, and new models of philanthropy require metrics for progress.

Additionally, there are other kinds of contributions that statisticians may be able to make. Some involve decision-theoretic aspects related to the philosophical foundation of human rights; others involve adjudication of grievances and balancing of competing interests. This chapter addresses a subset of those possibilities.

The members of the American Statistical Association are diverse in their backgrounds and broad in their thought. These new opportunities in the social and economic sphere provide the kind of challenge that our profession should welcome.

11.2 A Brief History

The evolution of human rights seems to move from the general to the specific. In the US, the Declaration of Independence suggested some basic principles: life, liberty, and pursuit of happiness. This perspective got lightly codified in the Constitution, and then detailed more particularly in the Bill of Rights. Subsequent court decisions have created a large body of highly specific rights.

The US law applies fully only to citizens; foreign nationals and prisoners of war have different status. Such exceptions seem contrary to the initial broad vision in the Declaration of Independence, and that discomfort with narrowness was mirrored in philosophical writings by Jeremy Bentham and John Stuart Mill. The slow evolution of their thinking, drawing upon Hume, Locke, and Rousseau, combined with the world's reactions to the horrors of the Holocaust and a burgeoning internationalism, eventually crystallized in the Universal Declaration of Human Rights (UN General Assembly Resolution 217 A (III), 1948). This document declared that citizens in all countries that were signatory to the United Nations signatory countries enjoyed common freedoms. Among the 30 enumerated freedoms were political freedoms, such as the following:

Article 13

1. Everyone has the right to freedom of movement and residence within the borders of each state.
2. Everyone has the right to leave any country, including his own, and to return to his country.

Article 15

1. Everyone has the right to a nationality.
2. No one shall be arbitrarily deprived of his nationality nor denied the right to change his nationality.

There were civil freedoms, such as the following:

Article 4

No one shall be held in slavery or servitude; slavery and the slave trade shall be prohibited in all their forms.

Article 5

No one shall be subjected to torture or to cruel, inhuman or degrading treatment or punishment.

Article 20

1. Everyone has the right to freedom of peaceful assembly and association.
2. No one may be compelled to belong to an association.

There were economic freedoms, including the following:

Article 17

1. Everyone has the right to own property alone as well as in association with others.
2. No one shall be arbitrarily deprived of his property.

Article 23

1. Everyone has the right to work, to free choice of employment, to just and favorable conditions of work and to protection against unemployment.
2. Everyone, without any discrimination, has the right to equal pay for equal work.
3. Everyone who works has the right to just and favorable remuneration ensuring for himself and his family an existence worthy of human dignity, and supplemented, if necessary, by other means of social protection.
4. Everyone has the right to form and to join trade unions for the protection of his interests.

And there were social freedoms, such as the following:

Article 1

All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of brotherhood.

Article 6

Everyone has the right to recognition everywhere as a person before the law.

Article 25

1. Everyone has the right to a standard of living adequate for the health and well-being of himself and of his family, including food, clothing, housing and medical care and necessary social services, and the right to security in the event of unemployment, sickness, disability, widowhood, old age or other lack of livelihood in circumstances beyond his control.
2. Motherhood and childhood are entitled to special care and assistance. All children, whether born in or out of wedlock, shall enjoy the same social protection.

Article 26

1. Everyone has the right to education. Education shall be free, at least in the elementary and fundamental stages. Elementary education shall be compulsory. Technical and professional education shall be made generally available and higher education shall be equally accessible to all on the basis of merit.

2. Education shall be directed to the full development of the human personality and to the strengthening of respect for human rights and fundamental freedoms. It shall promote understanding, tolerance and friendship among all nations, racial or religious groups, and shall further the activities of the United Nations for the maintenance of peace.
3. Parents have a prior right to choose the kind of education that shall be given to their children.

Note that in many examples the wording is vague (e.g., Articles 1, 6, and 15). And even when the intent is clear, as with the insistence on the equality of women, this agreement has had little force in most countries.

Western democracies frequently criticized countries that failed to meet these principles, and were criticized in turn for their own failures. During the cold war, the US emphasized political and civil rights; Communist nations such as Cuba and the U.S.S.R. were chided on Articles 13, 15, and 20, while Iran and Vietnam were taken to task on Articles 5 and 20, among others. Reciprocally, the United States and other wealthy democracies were condemned for minimizing social and economic rights, such as are found in Articles 23, 25, and 26. Communist nations felt their ethos was more responsive to economic security and social equality.

The ASA Committee on Scientific Freedom and Human Rights was founded in 1979, during this cold war climate, and its context was framed by human rights efforts led by President Carter, Amnesty International, and Human Rights Watch. The Committee's charge stressed case work on behalf of statisticians. It held letter-writing campaigns to support statisticians and allied scientists who were imprisoned, or denied access to foreign conferences, or who were prevented from emigration. Necessarily, the clientele were highly educated people, and thus tended to live in or be from developed countries. Thus, a great deal of the early attention was upon Soviet statisticians, and the Committee's efforts were congruent with the prevailing emphasis of Western democracies upon political and civil freedoms.

Over time, this changed. The casework began to include professionals in Africa, South America, and the Far East. The axis of discussion moved away from Communism, and the Committee members sought to define a research role for statisticians in human rights measurement. This led to a fresh view of the role of statistics, and there were long debates on what fell within the purview of the Committee. Questions that arose included whether the systematic exclusion of women from mathematics education was within the scope of the charge; whether providing the public with official estimates of the civilian cost of war was a scientific freedom issue; and whether clinical trials conducted in countries where life was cheap posed an abuse of rights.

This evolution led to one of the great research successes of the Committee. Several members and fellow-travelers developed theory and applications for multiple systems estimates of civilian casualties. These methods combined information from different human rights monitors, gravesite censuses, or other records to create a reconciled estimate of incidents. This enabled estimates and uncertainties for the total number of deaths and the total number of relocations. This evidence from Kosovo was used in the Hague during the trial of Slobodan Milošević, and it has

also supported investigations in Guatemala and East Timor. This new direction is one of several that relate to social and economic concepts of human rights. There are new opportunities for statistical research, and perhaps even for public influence. The remainder of the chapter lays out some of these topics.

11.3 Agent-Based Theories of Justice

The Veil of Ignorance is a Gedanken experiment introduced by the philosopher John Rawls (1971). It imagines that if one wants to decide which of two social systems is more just, one pretends that one's identity is going to be randomly assigned among all the citizens. Thus one cannot predict whether one will be male or female, straight or gay, black or white, rich or poor, sane or mentally ill, healthy or sick, young or old. A sufficiently impartial thinker would then have an objective criterion to compare the societies in terms of the expected quality of life.

Rawls hoped that this device would enable theorists displace themselves from ideology in making cultural judgments. He also hoped that this approach would undermine Utilitarianism, which puts no bound on the degree to which some people's interests can be subordinated to the advantage of the majority, where the advantage is calculated vaguely as some kind of expectation. Therefore, Utilitarianism might permit slavery, provided that the slaves are sufficiently few and that their work for the rest is sufficiently valuable.

From a statistical standpoint, Rawls' approach raises some questions:

- Rawls felt that the exchangeability of identities created a principle (the "Liberty Principle") that a fair social contract must provide minimum guarantees to the least well-off. But a gambler would disagree; the framers of the social contract might risk slavery if the odds were good.
- Rawls deduced that people should have equal assets, with the exception (the "Difference Principle") that some inequality was allowed if it enhanced the welfare of all (e.g., a doctor could make more money than others, since medicine helps everyone). But this appears to permit significant differences in compensation, perhaps comparable to those found in raw capitalism.
- Rawls maintained that this construct led to societies that maximized liberty, creating the largest possible benefit for all. But this is not obvious—perhaps the society is a local mode, and a different set of principles achieves a larger maximum.

Rawls was a philosopher, not a statistician. And we are statisticians, not philosophers, so in the spirit of exchangeability we cannot criticize this work too much. It has been a hotly debated perspective for years, and the arguments have grown more subtle and refined than the synopsis here suggests. See Wellbank Snook, and Mason (1982) or Baynes (1992) for more details.

Instead of constructing principles from this exchangeability, perhaps a more experimental approach is possible. A number of economists have used agent-based simulations to study the evolution of market behavior (cf. Jackson and Watts 2002,

Sallans et al. 2003). Epstein and Axtell (1996) have shown that simple agent rules can create complex behaviors that mimic societies. We believe that this approach can be applied to Rawls' formulation. One could devise agent-based models in which alternative rule sets for model societies play out, and examine the distribution in the quality of life among these.

More specifically, assume that the agents are initially placed at random in a p -dimensional space of latent psychosocial variables. These variables might be interpreted as, say, tolerance, criminality, political activism, and so forth. Each actor would then form friendship links (or enmity links) with other actors according to a probabilistic mechanism. Hoff et al. (2002) propose

$$\log \left(\frac{p_{ij}}{1 - p_{ij}} \right) = \mu_0 + \alpha_i + \alpha_j + \vec{\beta}^T \vec{x}_{ij} - d(\vec{z}_i - \vec{z}_j)$$

where p_{ij} is the probability of forming a link from agent i to agent j , α_i is the propensity for agent i to send a link, α_j is the propensity for agent j to receive a link, and $\vec{\beta}$ weights interaction factors between agents i and j that are captured in the vector \vec{x}_{ij} (e.g., the first component of \vec{x}_{ij} might equal one if both agents have the same ethnicity, and be zero otherwise; the second might measure their degree of kinship; the third might measure their difference in social class; and so forth). The \vec{z}_i and \vec{z}_j are the locations of the agents in the latent space.

At the first time step, suppose all of the actors form links according this model. Those who are close in the latent space are more likely to become friends than those who are distant, since increasing distance lowers the logit value. At the second time step, the rule is that each agent moves toward the mean of their friends. Then the process repeats. Similar rules would enable agents to disconnect friendship links, with probabilities that depend on distance and history.

Realistic implementation would require a more complex set of rules; most people drop friends less quickly than they add them, and most people only have time for a certain number of close friendships. If a similar model describes enmity ties, then with a little tuning one could get stable groups to form, and to move away from each other in the psychosocial space. Some agents might be highly charismatic, with large values of $\vec{\alpha}_i$; if they are in the intolerant region of space, they might attract a large group. And when there is an isolated group in a quadrant that is intolerant, criminal, and activist, then one has the social ingredients for oppression.

One nice aspect of the agent-based simulation is that one can allow time-varying covariates, to capture the effects of changing status, education, or aging (e.g., older agents can be made less likely to move toward criminality, and slower to change in general). Another nice feature is that it can give each agent a slightly different "personality" and thus alter the probabilities of certain kinds of interactions. A third nice feature is that some covariates may depend upon the location in space. Finally, a fourth feature is that different agents can value different goals; this leads to a more complicated model for how the agents move through the latent space. But a key problem is to develop a sufficiently rich class of probability models to describe this complexity.

11.3.1 Hierarchical Probability Models

The agent model is built around social factors that affect human rights violations. One key element is to divide the population into relatively homogeneous subpopulations which have relatively similar behaviors (or rule sets). These subpopulations would surely reflect social class, gender, and perhaps religion or race, depending upon the application of interest. Even so, it is desirable to allow variability among the members of each subpopulation.

Let D be the domain representing the possible states of an individual; this includes the location in the social space, as well as history and other covariates. As discussed previously, each simulated agent moves randomly in D according to agent-dependent, time-varying transition probabilities. Thus

$$P_a(t) = P_a \left[\vec{x}(t) | \vec{X}(s), s \in (0, t) \right]$$

is the probability distribution for agent a at time t on domain D given the entire history of agent a and all other agents up to, but not including, time t .

Ideally, one would like to create a full model of the behavior of every agent. However, such a modeling effort entails estimation of far too many parameters relative to available data.

Nonetheless, variability among agents is essential. As a compromise, some agent-based simulations might employ Bayesian hierarchical models. Specifically, one defines a manageably small number of subpopulations of individuals representing different socio-demographic and personality characteristics; e.g., one might have a subpopulation of lower-class white males and then allow agents within that group to have a range of incomes, employment status, and attitudes toward racism, according to some plausible hierarchical model. Individuals within a subpopulation have similar forms for their transition probabilities $P_a(t)$, but will evolve along trajectories that are more different than just random evolution would suggest because their initial covariates are also randomly different.

In many applications the domain D is taken to be discrete; this can be done by coarsening the continuous space in reasonable ways (e.g., counting income to the nearest \$1,000, or having a ten-point scale for attitudes on racism). Similarly, one can coarsen the time domain so that transitions occur at discrete (integer) times. This simplification allows much faster simulation, and there is a natural class of Dirichlet models for the transition probabilities.

To see some of the advantages of this discretization, ignore the full-time dependence and focus upon the transition matrix $P_a \left[\vec{x}(t) | \vec{X}(t-1) \right]$; its entries are the probabilities that agent a moves to state $\vec{x}(t)$ when the ensemble of agents is at $\vec{X}(t-1)$ in the previous time step, for all possible values of the discretized domain. Under the simplified model, this matrix can be chosen randomly from a distribution over all transition matrices. For both modeling flexibility and technical reasons, it is natural to use class-dependent Dirichlet prior distributions for each row of the

transition matrix. That is, with $\vec{X}(t - 1)$ fixed, the transition probabilities for $\vec{x}(t)$ are sampled from the Dirichlet distribution

$$f(\vec{p}|\vec{\theta}) = c \sum_{i=1}^{k-1} p_i^{\theta_i}$$

where c is a normalizing constant, k is the number of states in the discretized domain D , \vec{p} is the row of the transition matrix, and $\vec{\theta}$ contains the parameters that depend upon the past history of all of the agents.

Rapid calculation is possible because the Dirichlet is the conjugate prior for the multinomial; intuitive modeling is possible because researchers have experience with how the $\vec{\theta}$ terms carry the information.

An extreme alternative to this hierarchical approach is to assume that all individuals in a class have exactly the same transition probabilities. This alternative does not reduce the number of parameters to be estimated, but does reduce significantly the richness of the simulation.

The hierarchical model for discretized transition probabilities requires $k \times T \times n$ parameters, where k is the number of states, T is the number of time steps, and n is the number of subpopulations. This grows rapidly. But there are strategies that make this manageable; one can make assumptions that some states have probabilities that do not depend upon other actors, or that some states have probabilities that depend upon covariates in a simple way, or that some states have probability zero for given values $\vec{X}(t - 1)$. All of these assumptions are plausible for the kinds of human rights simulations under discussion.

11.3.2 Simulation

In order to understand the qualitative features of agent-based models of human rights behavior vis-à-vis clique formation, hierarchies, polarization, disenfranchisement, and so forth, one needs to run the model many times and summarize the results. The following are the principal steps in such a study:

1. For each individual agent a , assign its subpopulation. This may be done either deterministically or stochastically.
2. For each agent in a specified subpopulation, stochastically generate its initial location in psychosocial space, its covariates, and perhaps an initial set of social network links.
3. At each time step, agents form ties (of friendship, enmity, political alliance, religious affiliation, or all of these) according to an appropriate generalization of the model in Hoff et al. (2002). Ties from previous steps are unlikely to be broken or changed, but new ties may still form; the instantiation of this kind of complex behavior in agent rules is more of an art than a science and would require experimentation and judgment.

4. Between the time steps, all agents move in the psychosocial space toward those agents with whom they have positive links, and away from agents with whom they have negative links. (More complex rules for movement might ensure that cliques never degenerate into points, that rivalry enters into close partnerships, and so forth.)
5. Also between time steps, all agents move in the covariate subdomain within D (i.e., that portion of the domain that is not the psychosocial space) according to transition probabilities derived from hierarchical models.
6. Repeat steps 3 through 5, and track the behavior of interest (such as the formation of an oppressed minority).

The computational effort is, to first approximation, linear in the number of agents and quadratic in the number of states in the discretized domain D . The effort is also linear, obviously, in the number of replications of the simulation.

From a practical standpoint, one would start with small societies and simple rules for the agents. The hierarchical model for transition probabilities can also be made more transparent by making additional convenient assumptions about how probabilities are affected by covariates. For example, one might consider discrete choice models in which an agent moves at time t from state $\vec{x}(t-1)$ to one of a number of possible states according to a multinomial logistic model with probabilities determined by history and covariates. This leads to transition probabilities of the form

$$P_a \left[\vec{x}_j(t) | \vec{X}(t-1) \right] = \frac{\exp[\beta' x_j(t)]}{\sum_{i=1}^k \exp[\beta' x_i(t)]}$$

where the $\vec{\beta}$ is a vector of weights on the covariates (and recall that k is the number of possible states). This may oversimplify, but it seems flexible enough to capture many qualitative features of rights behavior; it also allows one to incorporate categorical features, such as race or gender, in a natural way.

11.4 Investment

Many modern philanthropies want to kick-start sustainable growth, rather than making donations that can only prolong intractable situations of appalling misery. This leads to the possibility of comparative cost-benefit analysis of social programs.

One-off cost-benefit analysis can be straightforward. For pennies a day one can vaccinate children. A simple calculation makes it clear that the benefit outweighs the cost. But a more complicated situation arises when one can invest in feeding children, or educating them, or vaccinating them. This makes it into a portfolio problem, and the prudent investor must do more mathematics to decide what is best. To make things more difficult, one often has much longer lists of choices, with complex interactions between them (educating hungry children is not very efficient—they do not learn as well). Also, one has to decide the time frame over which the analysis is done: for a few months, food is probably best; for a decade, vaccination is best; and

in the long run, the multiplier effect of education probably gives the largest return on investment.

Some economists (Sachs 2005) argue that it is possible to end poverty within 20 years. The world's wealth and technology can surely make large improvements, if resources are wisely directed. But there are statistical issues in how this can be achieved. One strategy is for many small philanthropies to cooperate in small-scale designed experiments to determine the effect of different kinds of interventions in different kinds of situations. For example, one might compare education versus infrastructure in countries that have high and low rates of corruption; this simple 2^2 experiment could give guidance on appropriate investment.

Similarly, economists use input–output models (mostly systems of linear equations) to model how resources flow through economic sectors. These models are poorly validated, but could provide important insight into the kinds of interventions that are most beneficial to developing nations. Statistical approaches to model validation are common in engineering; with some thought, data from simple experiments could help pin down input–output models for the economies of developing nations.

In this same spirit of smart philanthropic investment, we analyze the long-standing question of how much money an individual should contribute to charity. Many religions encourage their members to tithe, but this is an approximate approach to a moral problem that deserves serious attention.

Consider the following approach to determining appropriate altruism. First, identify the total need in the world. If there are I nations, suppose the distribution of income in nation i is $f_i(x)$ where x is annual income. Let t_i be the minimum annual income needed to live with dignity in that country. Then the amount needed in that nation is

$$V_i = \int_0^{t_i} (t_i - x) f_i(x) dx,$$

and so the total need in the world is $V = \sum_{i=1}^I V_i$.

Next, we figure out how much those with enough should give to those in want. To do this, first consider the utility function of money for the i th nation. Denote this by $U_i(x, y)$ where x is the amount of money one has, y is an additional amount of money, and $U_i(x, y)$ is how much that additional money is valued. A pauper would value one additional dollar at a dollar, but as wealth increases the additional dollar means less. There is a large body of experimental economics and psychology (cf. Rabin 2000, Kahneman 2003) that charts how this utility function works, and it works differently for different people. For our purposes, we assume that all people in a given nation have the same utility function, and we note that this assumption is reasonable but invites discussion.

Using Rawls's perspective, we assert the principle that the burden of altruism is distributed fairly when everyone who contributes should feel the same pain, as measured by their nation's utility function. A citizen of nation i with x dollars of

annual income who gives $g(x)$ to help the poor feels the pinch as $U_i(x - g(x), g(x))$. So set this to c , solve for $g(x)$ and one has the basis for determining what it is right to give.

To illustrate this, suppose that everyone values each dollar the same, so that $U_i(x - y, y) = y$. In this case everyone should contribute the same dollar amount, and the amount that is given for a fixed utility value c is

$$W_i(c) = \int_{t_i+c}^{\infty} c f_i(x) dx.$$

A different utility function model says that people who donate the same proportion of their income feel the same regret. This is a flat tax rate scenario, and the utility function is $U_i(x - cx, cx) = c$. In this case the amount that is given for a fixed utility c is

$$W_i(c) = \int_{(1+c)t_i}^{\infty} cx f_i(x) dx.$$

But experiments suggest a more complicated form for the utility function. Very wealthy people seem willing to tolerate giving a larger percentage of their wealth, and this is partially reflected in the graduated tax scales used in many countries.

From our standpoint, all that is needed is a utility function that is broadly accepted as sane by the standards of the nation. For that function, the amount of capital that is raised is $W_i(c)$. To end world poverty, one just solves

$$V = \sum_{i=1}^I W_i(c)$$

for c . This determines how much each person should give, in order that all feel the same inconvenience. The equation is easy to solve numerically, since each $W_i(c)$ is monotone in c .

11.5 Discounting Grudges

What is an appropriate model for discounting the past? During the disturbances in Kosovo, both sides pointed to massacres that the other had committed in the fifteenth century. And Sein Fein points to English abuses in Ireland stretching back Henry VII. To an outsider, it seems psychotic if politicians reach back 500 years to inflame their people. But jurists may see this as a continuum. The heritage of African-American slavery still lingers, and one can make a reasonable argument that compensation is due. There is still living memory of the genocide of Armenians

by the Young Turks in 1915–1916, but it is slipping rapidly off the table of active history. Many Palestinians living today hold valid legal deeds to property in downtown Jerusalem—how angry do they deserve to be, and what recompense is right? More topically, how can an international court calculate what a country owes when it is ultimately judged to have waged an unjust war? The Treaty of Versailles levied a penalty of £6,600 million against Germany at the end of World War I, but the Statistician John Maynard Keynes held that the penalty was too high and made World War II inevitable. There are some relevant data points.

The German government pays compensation to survivors of the Holocaust. The United States is paying the survivors of the Japanese internment camps. And recently U.S. District Judge Charles R. Norgle ruled against payment of reparations to African-Americans for slavery, on the grounds that no one directly affected is still alive. Each of these decisions represents some kind of negotiated agreement that, fairly or unfairly, balances the rights of victims against the assets of government authority.

These data could be used to fit curves for the degree to which claims get discounted with time. In practice, we expect that the estimated function will be extremely noisy, in part because there is little data and in part because different countries and courts have different attitudes. Of course, the calculation should be made in terms of standard dollars, to correct for changes in value over time.

A different approach to the problem of discounting past injuries lies in agent-based models (again). Axelrod (1984) created a competition he called the “Iterated Prisoner’s Dilemma” in which large numbers of agents competed against each other, over and over again. The simulated competition led to surprising insights about cooperation and may enable estimation of the cost of holding a grudge.

The Prisoner’s Dilemma is a game theory problem that took its name from the following hypothetical situation.

Suppose Abelard and Heloise are suspected of having jointly committed a crime, but there is insufficient evidence to convict. The police isolate them both, and separately offer each a deal.

- If Abelard confesses but Heloise does not, then he testifies against her and plea bargains down to no time while she gets 20 years.
- If Abelard confesses and Heloise does too, then both get 10-year sentences.
- If Abelard does not confess and neither does Heloise, then both get sentences of 5 years on trumped-up charges
- If Abelard does not confess, but Heloise does, he is sentenced to 20 years.

Abelard’s dilemma is whether to rat on Heloise. In game-theoretic terms, his payoff matrix is:

	Heloise confesses	Heloise does not confess
Abelard confesses	–10	0
Abelard does not	–20	–5

and an analogous table holds for Heloise. The minimax solution to this game is for both parties to confess. But this does not account for the possibility of repeated play.

Axelrod's competition allowed agents to have different rules on whether to confess or stand silent (defect or cooperate). The agents played each other repeatedly, choosing their partner at random from the pool of all partners. Each agent got a payoff according to the table above (actually, Axelrod added 20 points to each cell, since it was more convenient to talk about gains than losses, but this does not change the structure of the problem). The rules for the agents were generated by 63 experts in decision theory, psychology, and economics.

The winning rule set was submitted by Anatol Rapaport, a psychologist at the University of Toronto. His rule has been called "Tit-for-Tat" and specifies that the agent should start by cooperating, and thereafter do whatever the partner agent did in their previous round of play. Thus, if Rapaport's agent plays a copy of itself, it always cooperates (does not rat) and gains 15 points. If it plays an agent that defected (ratted) the last time it played, then Rapaport's agent will defect and gain just 10 points. Rapaport's rule set was the shortest of all entries, just five FORTRAN lines, and it won decisively.

Axelrod published the results of the competition, and invited all 63 experts to resubmit. Rapaport did not change his code, but everyone else did—as before, Rapaport's agent was the clear winner. Axelrod was intrigued, and altered the simulation so that agents who were more successful made duplicate copies of themselves in the pool of agents. He found that if he created a pool of cooperators and seeded it with defectors, the defectors gradually failed and disappeared. And if he created a pool of defectors and seeded it with cooperators, then again the defectors gradually failed and disappeared. All of this is quite surprising, from the perspective of conventional minimax theory.

Axelrod's study prompts the provocative possibility that free and fair trade, the rule of law, and human rights may create a stable equilibrium. If so, that would be wonderful news for human rights workers. Perhaps more study can elucidate this issue. It also suggests that populations of defectors are at an unstable equilibrium, and that a few cooperators will lead to total change.

From this perspective, Rapaport's agents need to learn to forgive. If they are placed in a society of defectors, they will all defect too, and no improvement is possible. But if, over a span of time, they are able to set aside an old grudge, then the population of defectors will eventually get replaced by a population of cooperators, and the net payoff will steadily increase. Agent-based simulations of this kind can explore the speed with which Tit-for-Tat agents need to forgive in order to maximize the net gains in payoff. And this curve will show how much it costs to hold a grudge as a function of time.

11.6 Conclusions

Statistical thinking can address new problems in the evolving theory of social and economic rights. This chapter has explored some of the nonstandard things we can contribute, suggesting

- ways in which human rights behavior can be simulated, in the context of a theory of justice based upon exchangeable agents;
- a procedure for determining how to fairly distribute the burdens of charity among the well-to-do and wealthy; and
- a methodology for studying the costs of compensation as a function of time, and thus appropriately discounting past injuries.

Our approach to these topics should be taken lightly—neither of the authors is an expert in these areas. But we think the problems are important, and that statisticians have a unique perspective on some of these issues.

But we do not mean to minimize the other important, and more traditional, work that statisticians can do in this arena. It is natural for us to develop procedures for tracking trends over time in quality-of-life indices, poverty measures, and so forth. Similarly, we can analyze comparative data from developing nations to identify factors that accelerate or retard growth. For example, there is much current discussion over the effect that official corruption has on the efficacy of public and private charities, and this impact could probably be quantified. And there are longstanding debates on what kind of interventions do the most to help the poor—most agree that ultimately one wants to invest in education, but the immediate need may be for food or medicine or security. This kind of tradeoff can be formulated as an optimization problem under large uncertainties, so statisticians have a role to play.

And that is the main message. Statisticians have a role to play. We have done it before and are doing it now, and there is good reason to hope that continued effort will bring fresh successes.

References

- Axelrod R (1984) *The evolution of cooperation*. Basic Books, New York
- Baynes K (1992) *The normative grounds of social criticism: Kant, Rawls, Habermas*. SUNY Press, Albany
- Epstein JM, Axtell RL (1996) *Growing artificial societies*. MIT Press, Cambridge
- Hoff P, Raftery A, Handcock M (2002) Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97:1090–1098
- Jackson MO, Watts A (2002) The evolution of social and economic networks. *Journal of Economic Theory* 106:265–295
- Kahneman D (2003) A psychological perspective on economics. *American Economic Review Papers and Proceedings* 93:162–168
- Rabin M (2000) Diminishing marginal utility of wealth cannot explain risk aversion. In: Kahneman D, Tversky A (eds) *Choices, values, and frames*. Cambridge University Press, New York, pp 202–208
- Rawls JB (1971) *A theory of justice*. Harvard University Press, Cambridge
- Sachs J (2005) *The end of poverty: economic possibilities for our time*. Penguin Press, New York
- Sallans B, Pfister A, Karatzoglou A, Dorffner G (2003) Simulation and validation of an integrated markets model. *Journal of Artificial Societies and Social Simulation* 6:1–33
- Wellbank JH, Snook D, Mason DT (1982) *John Rawls and his critics: an annotated bibliography*. Garland Publishers, New York

Chapter 12

Statistics and the Millennium Development Goals

David J. Fitch, Paul Wassenich, Paul Fields, Fritz Scheuren, and Jana Asher

12.1 Structure of Chapter

This chapter begins with the present introduction (Section 12.1), written to provide a quick roadmap of what is to come. Section 12.2 sets the human rights challenge of extreme poverty at center stage, showing why a book on human rights needs a chapter on the Millennium Development Goals.

The next three sections (Sections 12.3–12.5) are primarily anecdotal and historical. Indeed, the incidents related may no longer be typical. Still, we have included them because they make clearer the magnitude of the measurement problems to be overcome. They are instances where an opportunity was lost and precious resources wasted.

There is a touch of sadness in the examples, even some bitterness. We warn you of this in advance. As you know, statistical objectivity is hard to maintain on human rights and our struggle to stay balanced may show in places. We have to find another way to proceed if there are to be proper statistical evaluations of efforts to achieve Millennium Goals. Not all of our past efforts at development should be prologue to what needs to come now.

Section 12.3 gives short, mostly country-by-country vignettes that illustrate how data collection issues can impede a proper evaluation. In Section 12.4 we deal with a case study of how statistical methods can go astray. We reserve to Section 12.5 examples of some of the success stories that we have found.

But we want to move beyond anecdotes. However heartening the instances of real, albeit isolated, successes are, we are not doing enough. We, unabashedly, advocate the need to systematically embrace statistical impact evaluation *all the time*. This is our topic in Section 12.6. Too many good intentions have been chasing too little poverty reduction for too long. Given the scarcity of resources available for development programs, we can ill afford to devote them to seemingly good ideas that fail to deliver. And discovery of such failure usually depends upon the scrutiny of statistics.

Given the great need for augmenting the base of empirical knowledge about the development effectiveness of aid interventions, the number of rigorous impact evaluations conducted is surprisingly low. In Section 12.7 we explore why this is so.

One of the ways to redress this evaluation imbalance (see Section 12.8) is for program sponsors to incorporate program evaluation into the funding they provide. Otherwise, it would be irrational for the recipients of funding to do their own program evaluation with what always turn out to be scarce dollars. This leads us to the key role of the funder or donor in program evaluations. In Section 12.9 several suggestions are offered of ways donor agencies can put incentives in place to encourage more impact evaluations to occur.

Most of the other chapters in this book involve research done by just a few statisticians, loosely connected but still operating generally in small groups. The evaluation efforts needed to address the Millennium Development Goals are of a different order of magnitude. Yet, there is still a role for an individual with the right statistical training to offer solidarity to those in need. So we conclude the chapter with a short section (Section 12.10) that offers a starter set of answers to the question: “How can statisticians help?”

For those who want to learn more about this very complex subject we offer, in addition to the usual references, two further resources at the end of the text: A set of web pages on the themes covered in the chapter and an extensive bibliography prepared on relevant evaluation literature that Sarah Polen and one of us (Scheuren) have been studying in support of the Millennium Challenge Compact in the Republic of Georgia.

As a last thought, the goal of eliminating extreme poverty may be unattainable. Still, in rising together to the Millennium Challenge, we can achieve far more than if we stand by and do nothing. In this vein we end (in Section 12.11) with a success story by one of us (Fitch) from the early days of scientific social program evaluation.

12.2 Eradicate Extreme Poverty and Hunger

Poverty in many parts of the world today is staggering. Living and working in North America or elsewhere in the developed world, it is hard to imagine how anyone can survive on one or two dollars a day; yet billions of people do. And then, millions do not.¹ They die, in effect, of poverty. They die from preventable or treatable diseases because they cannot pay for medicine that costs as little as a few dollars. They don't go to school because they can't afford a uniform. They eat only corn, rice, and beans with the occasional chicken—because that is what they can afford to cultivate on their plot of land and they have no other source of income to buy more food.

In 2000 the United Nations specified eight Millennium Development Goals to help the developing regions of the world to improve their citizens' welfare (United Nations 2006). The first of those goals is to eradicate extreme poverty and hunger. Jeffrey D. Sachs, an economist and Director of the UN Millennium Project, has stated that 1.1 billion of the world's 6.5 billion people get by on less than US\$1 per

¹ www.globalissues.org/TradeRelated/Facts.asp is a rich repository of stylized facts about poverty in the world. It has references to sources which have documented that 30,000 children under 5 years of age die each day around the world, and that half of the world's population, nearly 3 billion, lives on less than two dollars a day.

day (Sachs 2005). He estimates that doubling current funding levels to about \$160 billion per year from now until 2015 would cut in half the proportion of the world's population in extreme poverty—a worthy yet ambitious goal.

The problem of world poverty is so big and the need is so urgent, there are no resources to spare. Consequently, there is too much at stake to base billion-dollar decisions on anecdotal evidence. It is vitally essential to establish the “best practices” based on rational and objective examination of programs before, during, and after implementation. Program sponsors constantly need to ask: “Does this program really work? Show me the evidence!” Too often, this is not the case and interventions do not produce the expected or even the purported results.

So what does statistics have to do with the Millennium Development Goals and worldwide poverty reduction? The world needs to generate some clear empirical evidence about what really works in poverty reduction. Statisticians have key roles to play in that effort.

In spite of decades of concerted efforts to reduce poverty in poor countries, poverty persists. This is not for lack of good intentions by organizations like the World Bank, the International Monetary Fund (IMF), United States Agency for International Development (USAID), various Non-Governmental Agencies (NGOs), and many others. Rather, the disappointing results of poverty reduction activities over the past 50 years suggest that poverty reduction is a complicated and difficult business, and that even the experts lack sufficient reliable empirical evidence upon which to base their decisions and design effective interventions.

The bottom line is that the development community needs to get serious about learning lessons from experience. Development interventions are effectively experiments, since one can never perfectly predict their impacts. Donors and governments should engage in systematic rigorous impact evaluation of their efforts, to generate reliable information about what works and what doesn't in poverty reduction. Moreover, we need to know not just what works, but what works best. In any case, whether as statisticians or simply as human beings, we have our work cut out for us.

12.3 Country-by-Country Examples of Gaps in Data and Methods

If the new Millennium Development Goals are to be met, then improvements are going to be needed to better measure and monitor how resources are expended and results assessed (e.g., Halberstadt and Wood 2006). There are several elements to consider. While they seem obvious, when stated, they are hard to achieve. Four problems will be mentioned and then illustrated with examples in this section, taken from around the world

1. *Lack of sound data collection.* A common problem in some countries is that reliable data collection methods are not practiced, with the result that decision-makers lack access to meaningful statistics that would allow them to target their efforts. There might be plenty of numbers, even, but existing data are

meaningless due to flaws in their collection. This is often the result when available personnel lack the proper skills to collect, analyze, and interpret data.

2. *Lack of a baseline.* To measure change, a baseline must first be established. A common difficulty is that status at the beginning of a program is not documented. Sometimes evaluation was an afterthought and so essential data collection was not done at all. Sometimes the timeline to do a thorough evaluation was longer than the time it took to mount the program and so the “before measurements” were a mixture of before and after measurements.
3. *Lack of good statistical analysis.* Here the issues are more subtle and can often be exacerbated by the international agency that comes in to offer assistance. And, frankly, all of us are still learning how to implement meaningful evaluations. The importance of random assignment is still new in many settings and harder to do as a result but the importance of persisting despite this cannot be overemphasized. Despite many good words about the need for a control group, e.g., when it comes down to practice, program implementers often oppose them.
4. *Opposition to experimental evaluation.* Because the randomization needed for experimental evaluation is often opposed in developing countries, USAID and the World Bank support of experimental evaluation is weak. Neither the donors nor the recipients are eager to hear bad news about ongoing programs.

12.3.1 Africa

Health is a fundamental determinant of economic development (Haines and Casells 2004, Lee et al. 2004). Consequently, three Millennium Development Goals target improving health. In one African country, however, a review of the official statistics showed that more people died of AIDS than were reported to have died! Policies can be no better than the data used in setting those policies. Obviously, those data were unreliable due to measurement errors or perhaps had been manufactured to achieve some political objective.

Nigeria. The population of the largest city in Nigeria—Lagos—is exploding. Uwe Deichmann, of the National Center for Geographic Information and Analysis, saw a published news item saying the population of Lagos was five million. He knew who had been conducting the study to determine the population of Lagos, so he called to congratulate them on completing the study, only to find out the study had not been completed and never would be. Then he was told: “You remember when I picked you . . . up at the Lagos airport [when you were last here] . . . you remarked to me that the city had grown greatly and now looked as if it might have five million inhabitants. [Since] this was the best estimate Nigeria was likely to have, I have subsequently employed it when people have asked me the question” (Deichmann 1997). Without good data, other data will fill the vacuum.

Senegal. Disasters can disrupt and derail economic progress. There was a law on the books in Senegal for 12 years for government agencies to cooperate in emergency preparedness. Did passing a law have an effect? No evaluation could be done since there was no measurement of cooperation nor of preparedness.

Tanzania. Over 40 years ago in Tanzania the Peace Corp instituted a program to construct small dams to help ameliorate poaching, which is economically driven. The idea was that with a reliable water supply, poachers would find farming more attractive than poaching. Has the program had any effect? No one knows, because a program evaluation was never performed. However, the poaching problem is reportedly worse than ever (Ferris 2006).

Uganda. As the number of people increases, food needs increase. This can lead to increased conflict between people and animals. In Uganda, elephants are protected and so their population is increasing. Elephants also find the farmers' crops attractive. To measure the size of the problem, a survey was taken asking the farmers:

“How often do the elephants come?” “Every night!”

“How many elephants come?” “100”

“100 elephants come every night?” “Yep!”

The unreliability of such survey data makes them useless for sound decision-making, but they can be misused to support the preconceived intention of a proposed intervention program. Instead, actual measurements are needed. Numerous elephant deterrent methods have been proposed ranging from removing their protected status (and then allowing farmers to shoot them) to installing electric fences to digging 50-mile long trenches. All have been tried, at great cost; yet have failed to solve the problem. A new low-cost idea is to plant chili peppers around the crops. Actual controlled experiments are being conducted to see if chili peppers will keep the elephants out and provide the farmers with a second cash crop.

12.3.2 Asia

Nepal. The Organization for Economic Cooperation and Development publishes annually 100 indicators for the world's leading economies (OECD 2006). But the same indicators of progress are not useful for developing countries. In the world today, more than a billion people lack access to clean water and 2.5 billion do not have proper sanitation (Wall 2006). Kathmandu, the capital of Nepal, rather than saying, “a river runs through it,” can best be described by saying “an open sewer runs through it.” Development cannot progress until the populace has access to clean water. As an official statistic, “Proportion of the Population with Access to Clean Water (ACW)” would be a more important measure of development than the ubiquitously used Gross Domestic Product (GDP). For developing economies, ACW should be reported at least as often as GDP.

Samoa. Data analysis skills, program impact evaluation, and evidence-based decision-making could have led to more desirable outcomes in Samoa. For example, a statistical analysis by an outside consultant revealed a problem with tooth decay and a possible link to a “luxury tax” on toothbrushes. Also, an NGO decided to implement a program to eradicate lymphatic filariasis, despite the fact that there were only six cases of filariasis in the country (compared to 15,000 cases of diabetes). And donors built an orthopedic hospital without noting that there were no orthopedic surgeons in the country.

12.3.3 *Latin America*

Bolivia. Even when studies are performed, poor study design can nullify the results. Based on broad generalizations—malnutrition is a significant problem in Latin America and Bolivia has the highest rate in Latin America (25 to 35 percent of the population according to the World Health Organization)—a school lunch garden program was implemented. Was there evidence of an impact? The study compared a treatment group (with gardens) and a no treatment group (no gardens), but without testing to see if there was any evidence of malnutrition to begin with. A before-and-after study design was needed.

The design of the experiment is always paramount. How were the schools in the study chosen? By convenience—one of the researchers had been there previously. Was there random selection of the participants? No, so maybe only healthy people self-selected to participate. Was there control for confounding from other health and nutritional programs? No—and, potentially, not a minor oversight. Further, since a positive effect was not found, the results were interpreted based on the desired outcome rather than statistical and scientific thinking: “Since the study found no evidence the gardens work, then there is no evidence they don’t work, so we can say they do work!” Right?

Guatemala. Microfinance is always supported by heart-warming “success” stories, but is there any evidence that it helps lift people out of poverty? As we know, without a full understanding of statistical techniques, people easily assert that correlations found in observational studies equate to a finding of causation. In Guatemala an extensive cross-sectional retrospective survey of microfinance participants was conducted. Some nice stories resulted, but no solid substantiation of reduced poverty. Prospective longitudinal studies are needed followed by proper statistical analysis. More importantly, in any program evaluation study, the most fundamental question is, “What is the measure of success?” Do the people in the program have the same measure of success as the program administrators? For example, is their objective to get out of poverty (the administrator’s goal) or to just solve a specific current personal problem? In Guatemala the later appeared to be the case.

Jamaica. Speaking of official statistics, the murder rate is so high in Jamaica that it is reported daily on television during the nightly news along with the stock market, weather, and sports statistics. It is one number that is readily available, easily understood and a highly relevant measure of whether conditions are getting better or worse.

12.4 Valadez’s Lot Quality Assurance Sampling Method of Evaluation

Now we consider a case study in how statistical methods can go astray. Just because statistics are used, it does not follow that evaluation is correct.

Dodge and Romig published in a 1959 book a method for answering a question such as, “From an industrial lot sample, what is the maximum number of defectives that can be found and still conclude with reasonable confidence that the rate in the lot is not more than, say 20 percent?” Valadez, as reported in his 1991 book, found a paper by George Stroh of the Centers for Disease Control referring to this industry method and suggesting that it could be used to answer questions in international health such as “Is the child vaccination rate in a community sufficient—say 80 percent or more—for protecting children from an outbreak of measles?” From this Valadez developed what he calls Lot Quality Assurance Sampling, but it is not the Dodge and Romig method. It incorrectly answers questions such as the one on vaccination. Whether the original error is due to Stroh or Valadez we do not know, as the Stroh paper was not published.

Let us look at the Valadez method. Table 12.1 gives the probabilities from the binomial distribution used by both Valadez and in the industry method, based on a sample size of 28 with a population unvaccinated rate of 20 percent. The first column is the number of defectives, e.g., children unvaccinated. The second column gives the probability of finding in the sample exactly this number of unvaccinated children, assuming an infinite population. Cumulative probabilities, 0–28, in the third column are child risks—the probability of concluding that children are protected when they are not.

Suppose d^* is the largest number of unvaccinated where a 20 percent null hypothesis is rejected and depends on α , e.g., if $\alpha = .05$ then $d^* = 1$ and the child risk is .02. If a mistake in thinking led, in effect, to an α such that $d^* = 9$, the child risk is .96. Probabilities of the fourth column are provider risks—the probability of concluding that providers have failed to achieve 80 percent coverage when in fact they had.

Suppose d^{**} is the smallest number of unvaccinated where a 20 percent null hypothesis is rejected (this depends on α); e.g., in testing this second null hypothe-

Table 12.1 Probabilities and cumulative binomial probabilities for numbers of children unvaccinated in samples of 28, given that the population unvaccination rate is 20 percent, i.e., the vaccination rate is 80 percent

d	P_d	$\sum_{d=0}^{d^*} P_d$	$\sum_{d=28}^{d^{**}} P_d$
0	0.00	0.00	1.00
1	0.01	0.02	1.00
2	0.05	0.06	0.98
3	0.10	0.16	0.94
4	0.15	0.31	0.84
5	0.19	0.50	0.69
6	0.18	0.68	0.50
7	0.14	0.82	0.32
8	0.09	0.91	0.18
9	0.05	0.96	0.09
10	0.02	0.99	0.04
11	0.01	1.00	0.01
12	0.00	1.00	0.00

Source: Prepared by chapter authors.

sis, if $\alpha = .05$ then $d^{**} = 10$ and the provider risk is .04. From Table 12.1, we see that if one or none are found unvaccinated, a 20 percent null hypothesis is rejected and a 20 percent or less hypothesis accepted, with a two percent child risk—2 percent of the time the conclusion that the population rate is 80 percent, i.e., that children are protected, will be wrong. But Valdez does not use this industry method. He writes that with finding nine or fewer unvaccinated it can be claimed that the 80 percent rate has been achieved. Something is wrong, since we see that with nine unvaccinated the child risk is 96 percent, so his conclusion that community children are protected will be wrong 96 percent of the time.

Perhaps Valdez is taking account of provider risk considerations, but he gives no such reasoning in his book. When asked for his basis he responded that if the vaccination rate is 80 percent, then in a sample of 28 children nine or fewer will be found unvaccinated 96 percent of the time. Apparently we are to conclude, falsely, that if nine or fewer are unvaccinated, then the community rate is 80 percent.

In different words, with related by not identical meaning, we speculate that Valadez reasoned as follows: He poses the assumption that the community rate is 80 percent, and draws a sample of 28 children. Finding nine or fewer unvaccinated is consistent with his assumption. He concludes that his assumption is confirmed and children are protected with 80 percent of the community being vaccinated. If ten or more are unvaccinated, then his 80 percent assumption is not confirmed and he would likely say that additional vaccination efforts are needed. This flawed argument gives Valdez's solution, as we see from row nine and column four that an 80 percent null hypothesis would not be rejected. But again, finding nine unvaccinated, his conclusion will be wrong 96 percent of the time.

Our sense that this flawed argument is the basis for his protocol is reinforced by what he has trainees do (cf. www.coregroup.org/working_groups/lqas_train.html, Module 1, p. 27). They put 80 green marbles and 20 red ones in a sack, shook the sack, and drew a sample of, say, 28. They would almost always find nine or fewer reds. Presumably they are to think they have demonstrated that with finding nine or fewer unvaccinated the community rate is 80 percent. We thank Dr Roberto Molina for clearly seeing this. With nine unvaccinated a 20 percent unvaccinated rate null hypothesis would not be rejected and thus one might falsely conclude that providers had succeeded. (In fairness, we note that in our review of the final contractor reports, although many say they used Valdez's LQAS method, they do not seem to be claiming that if 13 of 19 children are vaccinated, then the 80 percent success rate has been proven.)

Setting aside the technical errors above, the problem may be more fundamental.² Perhaps the real issue is that (1) both type I and type II errors should be considered and (2) error rates should be balanced based on public health costs as well as other costs. There are two types of error in a hypothesis test and it appears that only one is being controlled.

² This insight into the fundamental problem here was provided by Susan Hinkins during her review of this material.

Traditionally, when only the type I error was set, the null hypothesis was the conservative status quo and the analyst required strong evidence to reject the null hypothesis. One usually had a large enough sample size to provide some reasonable power. But $n= 28$ is not very large, so the tradeoff between the two kinds of error requires explicit attention. One must decide which type of error is the most important and set the test accordingly.

Here the question to be answered is whether the immunization level is adequate or not. If the immunization rate is less than 80 percent, the effort is considered inadequate (equivalently if the failure rate is 20 percent or higher, the effort is considered inadequate.). Let p_f equal the proportion in the population that are not immunized (failure rate), and it is computationally convenient now to consider $n = 19$. Assume two cases.

12.4.1 Case 1

The null hypothesis is that the immunization rate is adequate, i.e., $H_0: p_f \leq .20$; $H_a: p_f > .20$.

Let x_f = number of failures observed in the sample of $n = 19$. If the null is rejected when $x_f \geq 7$ then the α -level is approximately .07 (and this corresponds to the rule that you accept the immunization rate as adequate if the number immunized is greater than or equal to 13). So the test has controlled the type I error rate where the type I error is to “declare the immunization effort inadequate when in fact it was adequate.”

From a public health and human rights perspective, this does not seem to be the right error to control. Between the two types of errors one could make, the error of “declaring that the immunization rate is adequate when it is not” is more serious. This is not controlled.

Table 12.2 shows the probabilities of the two types of error, using this rule, for a variety of population p_f rates.

If the true failure rate is as high as 50 percent then the test has a 92 percent chance of finding the immunization to be inadequate. But if the failure rate is 30 percent, the test has only a 33 percent chance of finding the results inadequate. Is this sufficient power?

Table 12.2 Case I, probabilities of error

True percentage of failure	Probability decide immunization rate is	
	Adequate	Inadequate
10%	0.998	0.002
20%	0.932	0.068
30%	0.666	0.334
50%	0.084	0.916
70%	0.001	0.999

Source: Prepared by chapter authors.

Table 12.3 Case II, probabilities

True percentage of failure	Probability decide immunization rate is	
	Adequate	Inadequate
10%	0.42	0.58
20%	0.08	0.92
30%	0.01	0.99
50%	0.00	+0.99+
70%	0.00	1.00

Source:Prepared by chapter authors.

12.4.2 Case 2

Suppose instead that the hypothesis test was set up assuming that the results were inadequate, i.e., with the null hypothesis that the failure rate was 20 percent or more. In other words $H_0: p_f \geq .20$; $H_a: p_f < .20$. Then an α -level test of .08 would reject the null hypothesis when $x_f \leq 1$ or in other words one would conclude that the immunization was inadequate if $x_f \geq 2$ while the Case I test concludes that the immunization was inadequate only if $x_f \geq 7$. This is quite a difference and the probabilities of error are also quite different.

Table 12.3 shows the probabilities of deciding the immunization is adequate/inadequate using this test. With this test, if the failure rate is as high as 30 percent, this test will find it inadequate. Clearly, this test has much greater power to detect inadequate immunization rates, paying the price of a much larger chance of incorrectly declaring inadequacy when in fact the failure rate is quite low.

Intuitively, one sees that a sample size of 19 is too small. But more generally, neither test provides a reasonable balance of the two types of errors. Case I may be a human rights issue as it appears to undervalue the cost of inadequate immunization. Policy-makers must determine what the relative risks and costs are and develop a test protocol that addresses both types of errors, not just one.

12.5 Some Good Stories Need Telling Too

In this section, we complete our discussion by telling four more illustrative examples. These are instances, and there are many more (just not enough), where good intentions, good planning, good data, and good methods all came together.

Dominican Republic. It is possible to conduct program evaluations and research correctly. A statistically sound study using a matched pairs approach with control and comparison was conducted to evaluate the relative effectiveness of public health interventions on the prevalence of parasitic diseases among the poorest communities in the Dominican Republic. The commonly used interventions suggested by intuition produced no impact, while the study identified other interventions that could produce results and did so at the lowest cost—these programs are both effective and cost efficient.

Grenada. Another example of how statistics can promote economic development comes from Grenada. With the intent of attracting more *scuba* divers as tourists, a program was proposed to acquire a hyperbaric chamber to treat decompression illness from diving accidents. However, the population at greatest risk was not the tourists, but rather the local fishermen diving commercially for lobsters. A statistical analysis of a carefully constructed study showed that instituting the program would have increased the prevalence of decompression illness among the local lobster divers with devastating economic consequences. Thoughtful and proper program pre-evaluation averted the problem.

Nigeria (again). One of the authors (Fitch) recounts below a story of corruption, integrity, children dying, and a very good USAID man, with a shoebox full of money. Here is how it goes, in Fitch's own words:

When I was in Ibadan, Nigeria, the market was run by women—probably still is—with a most powerful one in charge. Now she had a physician son, Koye, a professor at Lagos University and a founder of the Institute of Child Health there—like Morley's at London.

[Later] USAID sent me back to Nigeria to look at Koye's under-fives clinics supported with their money. Arriving in Lagos and reporting to the US embassy, I found that people there wanted me to turn around and go back home. They had cancelled my hotel reservations and, as we were to learn, hotel rooms were impossible to find without pull. Two powerful people in the health ministry didn't want publicity for Koye's work. Almost all health money was supposedly being put into medical schools and hospitals.

Koye was concerned for ordinary children, some hundreds of thousands who were not living to their 5th birthday, and for their mothers. We went to his house. He got us a place [to stay] and agreed to meet with us the next day. The embassy had not told him we were coming. No arrangements had been made. When we met, Koye said that he would be pleased for me to visit all eight clinics throughout the country, but that would be impossible, as it would take weeks to make arrangements for hotels and drivers. He did agree to a second meeting. When I showed him a stack of air tickets he capitulated. I knew Nigeria. I didn't need help in getting around.

I have high admiration for the USAID man who arranged for funding for Koye's work—delivering to him, at one time, a shoebox full of Nigerian money in order to keep the clinic project going. He must have pulled a few strings to get the funding for this most important work. USAID needs more such people, passionately concerned to see improvements and fewer mothers grieving over dead children.

Koye's clinic in Yaba was very popular. Politicians wanted to take control. He pounded the table and said, "God will damn those who seek their own gain, and let children die." Koye later became Prime Minister.

Mexico. Consider, now, what may be among the best evaluations undertaken in the development field. This is a series of rigorous evaluations of a conditional cash transfer program in Mexico, originally called Progresá, now Oportunidades.

These evaluations are exemplary in several respects. The first is in the use of randomized assignment to the program. Since the program did not have sufficient resources to reach its entire target population in the first year of implementation, the administrators decided that random assignment was the fairest way to determine which communities would initially benefit, and which would join when the program went full scale 2 years later. Randomization eliminated the concern over sample selection bias in subsequent evaluations and provided a ready-made counterfactual

against which to compare the progress of beneficiaries, since both treatment and comparison groups are drawn from identical distributions of eligible participants.³

Another exemplary aspect of these evaluations was the public availability of the data that were collected. Because the data were published on the Internet (<http://evaloportunidades.insp.mx>), many social scientists have effectively conducted “free” evaluations of the impacts of Progres/Oportunidades by using the data in their own research. Finally, the evidence of positive impacts that the evaluations found is widely credited with the survival and continuation of the program across a change in political regimes, albeit with a new name.

12.6 A General Solution?

Impact evaluation, or program evaluation, aims to estimate returns on investments in poverty reduction. Impact evaluation estimates the effects attributable to development interventions—primarily done by quantitative techniques. This does not mean that qualitative (e.g., anecdotal) research methods do not play a role. But it is worth drawing a distinction between quantitative and qualitative techniques, in order to see how they complement each other and what kinds of evaluative questions each is best suited to answer.

By quantitative methods we refer to statistical or econometric methods using quantitative data, and typically involving large samples. Quantitative techniques allow rigorous estimation of average program impacts, and sometimes even the distribution of program impacts. Much akin to clinical trials in medicine, whence the econometrics of evaluating social programs draws much of its inspiration, quantitative analysis relies on statistical models to provide unbiased estimates of program effects and then tests whether these estimates are statistically significant compared with no intervention, or even a common alternative intervention. Quantitative methods are best suited to measure whether and how well a program achieved its development objectives. They can also be used to test alternative designs of an intervention—preferably using randomized assignment to different treatment groups (Campbell 1971, 1988a, 1988b), although propensity scoring methods combined with longitudinal data collection also can work well (e.g., Lokshin and Yemtsov 2003).

Quantitative methods can be something of a “black box.” A treatment effect can be estimated, but the particular aspects of a program that contributed to the observed impact may not be clear, particularly in the case of more complex interventions. And insofar as context matters, even though quantitative results can often be generalized to the population that the sample represents, they may not necessarily apply in other situations. So lessons learned may be viewed with some skepticism until a number

³ Work by Heckman on selection bias should be noted. He received a Nobel Prize (in Economics) for his research. Arguably, selection bias is among the most important problems that statisticians have to face. The latest issue of *Statistica Sinica* has a compelling editorial on this (Ming 2007).

of studies in a variety of contexts yield consistent results. Quantitative studies can also take a long time (often years) to implement before results are available.⁴ Strategic, long-term planners (e.g., those allocating budgets across potential interventions) should find the results of quantitative evaluations relevant and helpful.

Qualitative methods refer to in-depth research on small numbers of observations, or even on one observation. Qualitative techniques involve interviews, focus groups, direct observation, etc. Qualitative methods are best suited to process evaluations, where the research questions are focused more on how interventions were delivered (or how they were received). They are also well suited to exploratory studies where open-ended questions may be needed to capture information which the researchers did not anticipate. Because it is difficult to generalize from results obtained from few observations, qualitative methods are not well suited to answer the question "Did the program work?" Still, program managers and implementers can find the results of qualitative evaluations relevant and informative.

Quantitative and qualitative methods complement each other well, and the best evaluations make sagacious use of both.⁵ For instance, qualitative techniques can be used to compare program implementation to its design, while quantitative methods are used to generate an estimate of the program's final impact. Both methods may even be used in an iterative fashion within a single evaluation to generate questions for the next stage of research and to explore questions that arose from the previous stage.

The concept of "data transparency" also seems to fit naturally into this framework. Data transparency (e.g., by the creation of public use files) helps development as it can reduce aid misappropriation and corruption. The Mexico Progresá program is of this sort but data transparency may deserve attention on its own. It is a broader concept than program evaluation. For example, the Extractive Industries Transparency Initiative is a great example of how statisticians can help to change the world. This initiative is not linked to any specific development program.⁶

On the importance of transparency in program design, the work by Ritva Reinikka must be cited, including the famous example of how transparency helped to reduce grant leakage in Uganda schooling (Reinikka 1999, Reinikka and Smith 2004). The World Bank's Public Expenditure Tracking Surveys would be another important case.⁷

Of course, there can be potentially adverse impacts of aid on development at the macro level. See the recent paper by Bräutigam and Knack (2004). IMF's September 2005 World Economic Outlook also has some material on this topic, "Building Institutions" (Lallet al. 2005).

⁴ Longitudinal studies can obviously take as long as the researchers are willing to carry them out, and some last decades, though results are generally publicized for each round of data collection and analysis.

⁵ Rao and Woolcock (2004) wrote a good handbook chapter on integrating qualitative and quantitative methods for impact evaluation.

⁶ www.eitransparency.org/section/abouteiti

⁷ The World Bank Series PREMnotes (www1.worldbank.org/prem/premnotes.cfm) is an excellent source for more details on issues like this.

12.7 Why so Few Impact Evaluations?

Given the great need for augmenting the base of empirical knowledge on the development effectiveness of aid interventions, the number of rigorous impact evaluations conducted is surprisingly low. The primary reasons for this can be grouped into two categories: political economy and methodology.

Pritchett (2002) develops a succinct political economy model which shows why, in theory, we should expect very few rigorous impact evaluations. Among the myriad potential interventions competing for development resources, each has its own group of advocates that support it. Without hard evidence regarding the final impacts of an intervention, advocates can predict impressive results based on their best guesses. In this context, the first group to conduct a rigorous evaluation is actually taking a risk. Even if they find impressive results, they may sound no better than the unsubstantiated claims made by their competitors, and the politicians who allocate resources may not be able to distinguish between claims based on solid empirical evidence and those based on looser criteria. Moreover, if the evaluation results turn out to show less-than-hoped-for impacts, they run even more risk of sounding less compelling than best-case claims made by competing advocates.

Furthermore, politicians operate within the relatively short time horizons of election cycles. An evaluation whose results will not be known until after they are possibly or certainly out of office may be of no interest to them, or they may even perceive it as jeopardizing their legacy. Making evaluations happen is therefore dependent on a coalition of the curious—development organization staff, NGOs, finance ministry bureaucrats, and others. To have an incentive to support serious evaluation, they should still be doing similar work in a few years time, to be able to apply the lessons learned, and they should work across sectors to some extent, so as not to jeopardize their career on evaluation results that show their primary work to be undeserving of further funding.

Limitations in evaluation methodology also help explain the paucity of evaluations done. In particular, there is a lack of rigorous methods for evaluating macro-level interventions. Social programs can be successfully studied—you have participants and, in many instances, non-participants that could potentially serve as a control sample; even community-level interventions have potential comparison communities. But researchers have not come up with convincing ways to construct a counterfactual⁸ of a policy change that affects an entire country. Thus, the methods for evaluating interventions like legal reforms and structural adjustment programs remain less developed than those for evaluating micro-level interventions.

A reasonable question to ask is, “Why do the recipients of funding not do program evaluation routinely?” Could it be they do not want to know the answer? Since

⁸ The concept of the counterfactual lies at the heart of impact evaluation. The counterfactual is the (unobservable) hypothetical state of program beneficiaries, had they not benefited from the program. That is to say, in the absence of the program, what would have happened to those who did ultimately benefit from it? Comparing the counterfactual against the observed state of nature yields the estimate of impacts attributable to the program.

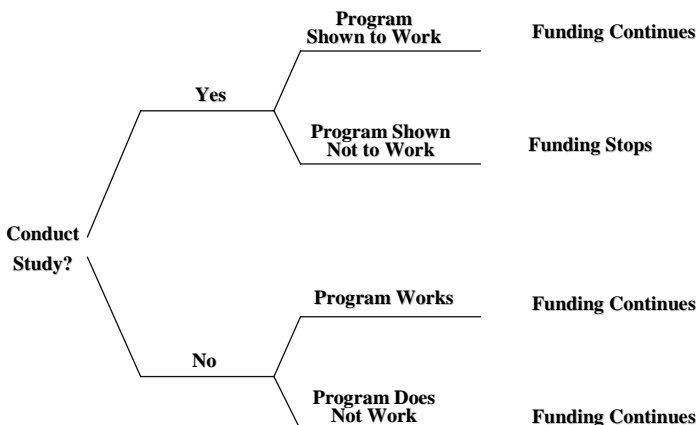


Fig. 12.1 Decision tree for Monitoring and Evaluation of Development Programs

funding continues whether the program is effective or not, but stops if it is shown not to be effective, it can be a rational decision on their part not to evaluate their programs rather than risk the termination of program funding. See the decision tree in Fig. 12.1.

It is therefore incumbent on program sponsors to incorporate program evaluation into the funding they provide because it is rational for the recipients of funding to not do program evaluation. This leads us to the key role of the funder or donor in program evaluations.

12.8 What Should Donors Be Doing?

When we look at the organizations that provide funding for development interventions, their track records on impact evaluation are mixed. In this section, we focus on two multilaterals, the World Bank and the Inter-American Development Bank, and two American agencies, the United States Agency for International Development (USAID) and a “new kid on the block,” the Millennium Challenge Corporation (MCC). Canadian, European, and UN counterparts to USAID exist but we are not informed enough about them to comment in detail. Nevertheless, the gists of our remarks probably apply broadly to other multilateral development organizations, and to the NGOs and government entities that implement development projects and programs.

12.8.1 World Bank

The World Bank seems to be the farthest ahead in terms of impact evaluation and has had a number of evaluation successes, many of which are worth imitating

(e.g., the already cited work of Lokshin and Yemtsov 2003). Yet arguably World Bank evaluations are too few, relative to their lending.

The World Bank has a fairly decentralized model for doing impact evaluations. The highest concentration in terms of impact evaluations per staff is probably in the Research Department, but many are done in other units. The Bank has a matrix organizational structure, and evaluations are carried out both in operations (where projects are designed and managed) and in other thematic units, which provide cross-regional technical and sector expertise to operational units.

It is perhaps a little surprising that the Independent Evaluation Group (IEG), formerly the Operations Evaluation Department (OED), performs few rigorous impact evaluations. Probably this is because they have a broad mandate and a fairly small budget. They conduct about one rigorous impact evaluation a year. By its own calculation, IEG (then OED) between 1980 and 2002 conducted 23 rigorous impact evaluations. See www.worldbank.org/ieg/docs/world_bank_oed_impact_evaluations.pdf for details.

This decentralization of impact evaluations tends to mean that individual projects are singled out for evaluation in a somewhat ad hoc fashion. This has the advantage that the person or team that decided to evaluate the program in question can often follow it over the course of several years as it is designed and implemented. This means that project and evaluation design can potentially be done simultaneously, which is necessary for most rigorous evaluation designs, since baseline data collection and random assignment to treatment groups cannot be done once a program has begun.

A drawback of the decentralization is that programs tend to be evaluated in isolation. This means that drawing up a body of evidence by comparing several evaluations of similar projects can take many years. Various agents might coincidentally choose to evaluate similar programs. Furthermore, the fact that the evaluations' designs may vary significantly can complicate the comparison of results across studies. Conducting a meta-analysis of the results, necessary for broader lesson learning, can thus be exceedingly difficult.

12.8.2 Inter-American Development Bank

The Inter-American Development Bank (IDB) has a mission similar to the World Bank, but operates exclusively in Latin America and the Caribbean. Like the World Bank, it has conducted and supported rigorous impact evaluations of its interventions but, as with the World Bank, this covers a similarly small percentage of its overall operations.

By contrast, the IDB has a relatively centralized model of conducting impact evaluations. While a few impact evaluations are contracted by operational units, the Office of Evaluation and Oversight has a fairly new mandate (as of 2004) to conduct several impact evaluations a year. They have chosen to organize these by selecting a few themes annually, and evaluating all the recent projects they can within each theme.

However, the Office's mandate also specifies that the impact evaluations be conducted *ex post*, after a project has closed. This makes applying rigorous impact evaluation methods a difficult proposition. The evaluation team must either reconstruct an estimate of baseline characteristics using existing data or rely on *ex post* cross-sectional analysis. Even using sophisticated methods, the latter is subject to inherent weaknesses in terms of convincing a skeptic that all potential sources of bias have been eliminated.

Nevertheless, the IDB has, in some cases, been successful at identifying sources of existing data that can be used for reasonably sound evaluations, including a few cases where the project design included a component for collecting data to be used in an impact evaluation. In an IDB evaluation cycle using this *ex post* approach, the steps are to first identify projects in the selected themes, eliminate the majority of projects which are "not evaluable" for lack of data, and do the best one can with some combination of existing and/or new data for the few projects that remain.

This more centralized, theme-based approach to selecting projects for evaluation allows a concentration of resources on evaluations most likely to generate reliable results. By contrast, in a context where a fixed budget is assigned to the evaluation of a particular project, one sometimes observes evaluations that are so resource-constrained that they have little hope of generating a reliable estimate of impacts.

12.8.3 USAID

USAID expresses support for quantitative evaluations and extensive data collection efforts, notably in the Demographic and Household Surveys performed in over 75 countries (www.measuredhs.com/). Our review of some 100 mid-term and final reports that their contractors have delivered (<http://dec.usaid.gov/>, Advanced Search) shows that before–after data are often collected. These include, e.g. the percent of (a) children under two who have received the recommended vaccinations, (b) women who know two ways of protecting themselves against AIDS, (c) mothers who can recognize pneumonia and know how to obtain antibiotics.

Nonetheless, our reading of the material available online suggests that most evaluations of USAID interventions are qualitative, and virtually never are quantitative techniques used to estimate counterfactuals. Therefore, they attribute all changes in beneficiary welfare to the intervention in question—clearly a statistically weak approach.

12.8.4 Millennium Challenge Corporation

Created in 2004, the Millennium Challenge Corporation is, like USAID, a United States government entity. But unlike USAID, which finances specific projects, MCC provides budgetary support to eligible countries with approved development plans and, indeed, plays an active role in framing projects and especially in developing

ingenious ways to evaluate them rigorously. MCC has, as one of its four key principles, a “focus on results.” This includes, as its website⁹ stipulates, that it has to have “a plan for effective monitoring and objective evaluation of results.”

It is often maintained that even though randomized trials are the “gold standard” of research, they are too difficult to conduct and, therefore, it is unrealistic to expect program analysts and policy-makers in the real world to do them. No doubt they are hard to do but not impossible and their definitive conclusions make them essential.

An example that invalidates the contention that they are unrealistic is the Poverty Action Lab at the Massachusetts Institute of Technology, one of MCC’s contractors. The Lab was founded in 2003 to use randomized trials to evaluate anti-poverty programs. Michael Kremer, who is affiliated with the lab said, “Development goes through a lot of fads. We need to have evidence on what works” (Minkel 2005). Evaluation studies at the Lab have ranged from microfinance programs to measures of public health to AIDS prevention and trails of the poor.

Random assignment to participation in a program is the most convincing evaluation method available (Duflou and Kremer 2004) and is being used, e.g., by MCC in the Republic of Georgia in an Agri-Business Development Activity (ADA) initiative that involves over \$15 million in grants to farmers. The first wave of random assignments was made in October 2006 and many more waves are to come (Scheuren et al. 2006).

Randomization can be used as a fair method of allocation when resources are not sufficient to reach all target beneficiaries, as with the ADA program in Georgia.¹⁰ It can, of course, be used to test alternative program designs when resources are sufficient for the entire target population to participate. Such an application of randomization is being contemplated in an MCC initiative in Benin (Felkner et al. 2006).

Early planning can vastly improve an impact evaluation. Randomization, for instance, is only possible when an evaluation is planned during the design phase of a program. But the use of propensity scoring methods cannot be an afterthought either. Early planning, while the exception more than the rule in most evaluation practice, is a basic tenet at MCC. Typically, the evaluation planning goes along in tandem with the planning for particular program support.

Public availability of data can also yield benefits by stimulating research on a variety of program impacts. MCC is planning to make microdata record files available as the evaluations are done. The World Bank has a regular habit of releasing files to interested researchers too. These have been a source that MCC used to try

⁹ www.mcc.gov/about_us/overview/index.shtml

¹⁰ The Chief Executive office of Millennium Challenge Georgia, the entity implementing promising poverty reduction program initiatives in Georgia, is Lasha Shanidze. Celeste Tarricone in the Monitoring and Evaluation Division at MCC in Washington works with Millennium Challenge Georgia staff, led by Mamuka Shatirishvili, to implement various impact evaluations of the compact. In Georgia, a combination of propensity scoring (primarily for a large road improvement initiative) and randomization (for a direct grant program) are being employed to carry out the evaluation. An extensive effort to review the evaluation literature both generally and in a Georgia-specific context was carried out by Sarah Polen and others and the references found are included in the bibliography that accompanies this chapter.

out its thinking on the evaluation of its development initiatives in Georgia, Armenia, and elsewhere.

Academics, if given data access, will often do valuable evaluation research at no additional cost. This can help generate deeper and broader knowledge of intended and unintended effects of the program. This information can, in turn, help politicians and others make better-informed decisions regarding possible changes to the program design or implementation mechanisms and to funding allocations.

12.9 Institutional Incentives Donors Could Implement

Donor agencies are in a position to put incentives in place to encourage more impact evaluations. One idea is to establish a funding mechanism that would provide matching grants to borrower governments who are willing to spend some of their own money on program evaluation. This would give governments an incentive to properly evaluate the programs they are borrowing to finance, and gets the donor and the borrower to share the cost of the evaluation. It may also offer the donor an entry point for proffering technical guidance when government staff lack expertise in evaluation methods.

Another step donors could take is to require that data used in evaluations be made publicly available (as MCC is planning on doing). Public availability of data could have two beneficial effects. One would be to allow for verification of results by other researchers, as well as extending evaluations to program effects that may not have been included in the initial evaluation, as seen in the case of Progresa. The other would be to set a precedent of public data availability to show that it need not jeopardize the confidentiality to which survey respondents are entitled. Governments and statistical institutes could see that releasing their data in appropriate formats need not compromise confidentiality, and perhaps become more inclined to share data.

Donors should commit to always publicize the results of evaluations—good, bad, or indifferent—to build up an unbiased body of evidence. Publication bias tends to mean that only studies that produce statistically significant results get published. This can produce a misleading “big picture” if studies that find zero or negative results are not as likely to be published or publicized as those which find positive results. Meta-evaluations will then produce skewed results, due to biased omission of some studies.

12.10 How Can Statisticians Help?

So what is the role for statisticians in particular in impact evaluation for poverty reduction? Statisticians seem well suited to fill several lacunae in the big picture. Two, which come readily to mind, are sample size and statistical power considerations, and new statistical methods development, especially for estimating

macro-level policies. Propensity scoring techniques, while common now, have assumptions that are hard to check and need more development at both a practical and theoretical level (e.g., Rosenbaum and Rubin 1983, 1985).

Sample size and statistical power are not often taken into consideration in program evaluations. In fact, these calculations are rarely done in economic evaluations. This may be partly due to the fact that doing a priori calculations of adequate sample size is complicated in most social science research designs. Also, economists and econometricians are not always well versed in doing these calculations.

So in practice, determining an adequate sample size in the planning phase of an evaluation is usually overlooked or deemed too complicated to carry out. This is a shame. "Running the numbers" can help to avoid wasting resources on studies that have little hope of detecting an effect of interest because they are underpowered. Even within the framework of a single evaluation, calculating sample size and power can help to focus researchers on questions they can reliably answer within their resource constraints.

Surprisingly, a clear exposition of how to go about determining sample size in the setting of a complex social science research design is difficult to find. In our view, it does not receive sufficient treatment in statistics or econometrics texts, or in journal articles.¹¹

The development of new statistical techniques to estimate program impacts could greatly improve the quality of evaluations and the reliability of their results. In particular, as mentioned above, techniques for evaluating the impacts of macro-level interventions have not been as thoroughly developed or refined as those for micro-level interventions. This hinges on the fact that micro-level projects have a natural comparison group in non-participants that can be used to construct an estimate of the counterfactual.¹² By contrast, since macro-level interventions affect the entire population, the evaluator cannot rely on non-participants to estimate a counterfactual.

Progress is being made in developing methods for evaluating the impacts of macro-level programs and policies,¹³ but they are perhaps still generally deemed to provide less-definitive results than some of the methods that can be brought to bear on micro-level projects. This is an area on the research frontier that could surely benefit from the attention of statisticians.

We can all strive to create more demand for evidence-based decision-making and for impact evaluations to inform public policy. This applies not only in the context of development and poverty reduction, but also for public policy in general. In this connection, there are many obvious roles for statisticians: (1) providing guidance in designing studies, (2) monitoring the execution of studies, (3) analyzing and interpreting the results, (4) advocating for meaningful official statistics as performance measures, and (5) lobbying for funding to support program evaluation, (6)

¹¹ The difficulty is in how one handles the integration of complex sampling structures, and problems like missing and misreported data. Since speculations about power are required at the design stage no general solution seems attainable.

¹² See Footnote 8 for the definition of counterfactual.

¹³ See, for instance, the second half of Bourguignon and Pereira da Silva (2003).

participating in development programs by training people in data collection and survey sampling and, where possible, (7) offering some of our efforts *pro bono*.

We have yet to see much in the way of applications where randomization and Propensity Score Matching (PSM) are used together; we view this as an area where more research is needed. PSM assumes that unobservables have a degree of stability over time that allows one to argue that their effects cancel. Only when randomization is used can this be tested. And, usually, the amount of randomization that occurs in any one study is insufficient for the detection of assumption failures. This is another reason for donors to sponsor meta-analysis efforts. And to push this point a bit further, it is a non-trivial task to use a study afterward in a meta-analysis, requiring the active participation of the original investigators in developing the needed metadata and paradata during program execution (Scheuren 2005). As we have said elsewhere, for an evaluation to be well conducted it must be built into the evaluation program from the beginning. For a meta-evaluation to be well done, it too must be designed from the beginning.

12.11 Experimental Evaluation Is Essential

This chapter has tried to make a compelling case that we need to see more experimental evaluation in the support of Millennium Development Goals. It is a direction needed for effective accountable poverty reduction. A little history might be in order to draw out the importance of this point one last time. As one of us (Fitch) tells it, in the mid of late 1970s a group, inspired by Donald Campbell's then unpublished paper "Methods for the Experimenting Society," worked to introduce experimental evaluation into efforts of the US Federal Government. The evaluation of a home care program in the state of Massachusetts could have been the first. As he said, "I don't suppose it would have happened without our efforts, since random assignment was powerfully opposed there, as is frequently the case today."

But how can such opposition be overcome so that poverty reduction can proceed at a good pace? One partial answer is for those whose discipline includes the design of research to say clearly what they know. If you want evidence as to which of two treatments is the more effective, you assign one treatment to a random set of areas and the other treatment to a second set of areas, randomly selected from the same population. Here statistics has a responsibility. To encourage the profession to think about the contributions we need to make, at the 2006 Joint Statistical Meetings the session "Statistics and the Millennium Development Goals" was organized. That was the genesis of the present chapter.

But having statisticians simply presenting the facts is not enough. In the case of the Massachusetts home care program, Fitch cites powerful opposition to experimental evaluation from the Visiting Nurses Association. The nurses felt that they should make the final decision as to which persons would receive an offer of the home care option. What was suggested, and supposedly occurred, was that the nurses would select a large enough group that twice as many were selected as

there were places. Then there would be random selection from within this group to establish treatment and control groups.

A related problem, Fitch continues, can be described as follows. “As is reasonable, there will be people in a poor country, like where I live in Guatemala that, seeing the lifestyle of people in the US, seek the same. They reason, with logic, that if they can find ways of directing some seemingly desirable experimental program to their district they will be more likely to be selected and thus receive money allowing a more-nearly-equal-north American life style. They seek ways of directing such programs to their districts, and hence not allowing a good evaluation to take place.”

So we have problems in *measurably* achieving the millennium development goals. As all the authors of this chapter have experienced, there will be powerful opposition to random assignment. Fitch concludes his story by using the phrase “at first” knowing what happened afterward. “The Massachusetts home care evaluation work began in the late 70s. When I had a chance in the late 80s to look at home care work in the intervening 10 years I found all of the evaluations had been experimental!” Jeffrey Sachs has said, “Let’s use real evidence and strip away the ideology” (Powell 2006). The authors of this chapter and, by now, most of its readers would agree. Let it be so!

12.12 Web Resources

- Campbell Collaboration: www.campbellcollaboration.org

The international Campbell Collaboration (C2) is a non-profit organization that aims to help people make well-informed decisions about the effects of interventions on the social, behavioral, and educational arenas. C2’s objectives are to prepare, maintain, and disseminate systematic reviews of studies of interventions. We acquire and promote access to information about trials of interventions. C2 builds summaries and electronic brochures of reviews and reports of trials for policy-makers, practitioners, researchers, and the public.

- OECD Development Assistance Committee Network on Development Evaluation: www.oecd.org/department/0,2688,en_2649_34435_1_1_1_1_1,00.html

The OECD’s DAC Network on Development Evaluation contains links to many online resources on impact evaluation, including evaluation offices within many multilateral and bilateral aid donors, and evaluation societies around the world. This site’s focus is on evaluation broadly speaking, and information on impact evaluation more strictly speaking can be hard to find.

- Center for Global Development: www.cgdev.org/section/initiatives/_active/evalgap

The Center for Global Development has an Evaluation Gap Working Group, whose purpose is to address the problem of the lack of knowledge about the effectiveness of social programs in low- and middle-income countries.

- Poverty Action Lab at MIT: www.povertyactionlab.com/

The Abdul Latif Jameel Poverty Action Lab serves as a focal point for development and poverty research based on randomized trials. The objective is to improve the effectiveness of poverty programs by providing policy-makers with clear scientific results that help shape successful policies to combat poverty. The Lab was started in June 2003 by Professors Abhijit Banerjee, Esther Duflo, and Sendhil Mullainathan at the Massachusetts Institute of Technology. The Lab was renamed in honor of Abdul Latif Jameel in October 2005.

- Poverty Facts and Stats: www.globalissues.org/TradeRelated/Facts.asp

This site contains an impressive array of well-documented stylized facts regarding poverty around the world and global income distribution. (Other pages on the site are not so well documented.)

- Millennium Development Goals: www.developmentgoals.org

This site has links to the list of MDGs, and to some information to track progress on those indicators for each country.

- World Bank impact evaluation page: www.worldbank.org/impactevaluation/

An impact evaluation assesses the changes in the well-being of individuals that can be attributed to a particular project, program, or policy. This website aims at disseminating information and providing resources for people and organizations working to assess and improve the effectiveness of interventions aimed at reducing poverty.

- World Bank World Development Indicators: www.worldbank.org/data/wdi2005/

References and Selected Evaluation Bibliography

- Bourguignon F, Pereira da Silva LA (eds) (2003) *The impact of economic policies on poverty and income distribution: evaluation techniques and tools*. World Bank and Oxford University Press, Washington DC
- Brütigam DA, Knack S (2004) Foreign aid, institutions, and governance in sub-Saharan Africa. *Economic Development and Cultural Change* 52:255–285
- Campbell DT (1971) *Methods for the experimenting society*. Unpublished; available from dfitch@uvg.edu.gt
- Campbell, DT (1988a) Reforms as experiments. In: Overman ES (ed) *Methodology and epistemology for social science: selected papers*, University of Chicago Press, Chicago
- Campbell, DT (1988b) *The experimenting society*. In: Overman ES (ed) *Methodology and epistemology for social science: selected papers*, University of Chicago Press, Chicago
- CIDA (2004) *CIDA evaluation guide*. Canadian International Development Agency, Ottawa ([www.acdi-cida.gc.ca/INET/IMAGES.NSF/vLUIImages/Performancereview5/\\$file/Evaluation%20Guide.pdf](http://www.acdi-cida.gc.ca/INET/IMAGES.NSF/vLUIImages/Performancereview5/$file/Evaluation%20Guide.pdf) accessed May 6, 2007)
- Deichmann U (1997) African population database documentation – population projections and data quality, additional sources of error. National Center for Geographic Information and Analysis, University of California, Santa Barbara (www.glowa-volta.de/cd_v3.1/landuse/populat/africa.htm, accessed May 6, 2007)
- Duflo E, Kremer M (2004) Use of randomization in the evaluation of development effectiveness. In: Feinstein O, Ingram GK, Pitman GK (eds) *Evaluating development effectiveness* (World Bank

- Series on Evaluation and Development, Volume 7), Transaction Publishers, New Brunswick, pp 205–232 (http://econ-www.mit.edu/faculty/download_pdf.php?id=759, accessed May 6, 2007)
- Duflo E (2006) Field experiments in development economics. Prepared for the 2005 World Congress of the Econometric Society, London (http://cemmap.ifs.org.uk/papers/vol2_chap13.pdf, accessed May 6, 2007)
- Felkner J, Kohagen K, Mark K, Reynolds M, Telgarsky J, Wolter, K (2006) Design report: impact evaluation design and implementation services – Benin. Report prepared for Millennium Challenge Corporation
- Ferris R (2006) Heartbreak on the Serengeti. *National Geographic* 209:2, 4, 7–12, 14–20, 22–26, 28–29
- Halberstadt S, Wood S (2006) Global public health: the new frontier of statistics. *Chance* 19:35–39
- Haines A, Cassels A (2004) Can the Millennium Development Goals be attained? *The British Medical Journal* 329:394–397 (<http://bmj.com/cgi/content/full/329/7462/394>, accessed May 6, 2007)
- Heckman J, Ichimura H, Todd P (1998) Matching as an econometric evaluation estimator: evidence from evaluating a job training program. *Review of Economic Studies* 64:605–654
- Heckman J, Ichimura H, Smith J, Todd P (1998) Characterizing selection bias using experimental Data. *Econometrica* 66:1017–1099
- Lall S, Spatafora N, Sommer M (2005) Building institutions. Chapter 3 in *IMF World Economic Outlook*, International Monetary Fund, Washington DC
- Lokshin M, Yemtsov R (2003) Evaluating the impact of infrastructure rehabilitation projects on household welfare in rural Georgia. World Bank Working Paper 3155. World Bank, Washington DC
- Lee K, Walt G, Haines A (2004) The challenge to improve global health. *Journal of the American Medical Association* 291:2636–2638 (<http://jama.ama-assn.org/cgi/content/full/291/21/2636>, accessed May 6, 2007)
- Minkel JR (2005) Trials for the poor. *Scientific American* 293:18–20
- Organization for Economic Cooperation and Development (2006) *OECD Factbook 2006 Economic, Environmental and Social Statistics*. OECD Publishing, Paris (<http://new.sourceoecd.org/factbook>, accessed May 6, 2007)
- Powell CS (2006) The discover interview: Jeffrey Sachs. *Discover*, November, pp 60–65
- Pritchett L (2002) It pays to be ignorant: a simple political economy of rigorous program evaluation. *Journal of Policy Reform* 5:251–269 (<http://taylorandfrancis.metapress.com/openurl.asp?genre=article&issn=1384-1289&volume=5&issue=4&spage=251>, accessed May 6, 2007)
- Rao V, Woolcock M (2004) Integrating qualitative and quantitative approaches in program evaluation. In: Bourguignon F, Pereira da Silva LA (eds) *The impact of economic policies on poverty and income distribution: evaluation techniques and tools*. World Bank and Oxford University Press, Washington DC (www.cultureandpublicaction.org/bijupdf/ch08.pdf, accessed May 6, 2007)
- Ravallion M (2001) The mystery of the vanishing benefits: an introduction to impact evaluation. *World Bank Economic Review* 15:115–140 (http://siteresources.worldbank.org/INTISPMA/Resources/383704-1130267506458/Mystery_Vanishing_Benefits.pdf, accessed May 6, 2007)
- Reinikka RS (1999) Using surveys for public sector reform. PREM Notes No. 23. World Bank, Washington DC
- Reinikka RS, Smith N (2004) Public expenditure tracking in education. International Institute for Education Planning, Peru
- Rosenbaum PR, Rubin DR (1985) Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1): 33–38.
- Sachs JD (2005) Can extreme poverty be eliminated? *Scientific American* 293:56–65
- Scheuren F (2005) Paradata and metadata. Proceedings of the 2005 Statistics Canada Methodology Conference (www.statcan.ca, accessed May 6, 2007)
- Scheuren F, Felkner J, Polen S (2006) Georgia summary design report. Prepared by the National Opinion Research Center for the Millennium Challenge Corporation
- Telgarsky J, Scheuren F (2006) Benin summary design report. Prepared by the National Opinion Research Center for the Millennium Challenge Corporation

United Nations (2006) Millennium project (www.unmillenniumproject.org, accessed May 6, 2007)
 Wall B (2006) Facing global challenges while turning a profit. *International Herald Tribune* July 8–9:15–16

Impact Evaluation: Overview

- Baker JL (2000) Evaluating the impact of development projects on poverty: a handbook for practitioners. World Bank, Washington DC
- Chase RS (2002) Supporting communities in transition: the impact of the Armenia social investment fund. *World Bank Economic Review*, 16:219–240
- Chen S, Ravallion M (2003) Hidden impact? Ex-post evaluation of an anti-poverty program. World Bank poverty research working paper 3049. World Bank, Washington DC
- Fields P (2006) Statisticians and the millennium development goals. Paper presented at the annual Joint Statistical Meetings, August 6–10
- Fitch DJ (2006) A role for experimental evaluation in efforts to achieve millennium development goals. Paper presented at the annual Joint Statistical Meetings, August 6–10
- Grosh M, Glewwe P (eds) (2000) Designing household survey questionnaires for developing countries: lessons from 15 years of the living standards measurement study, Volumes 1–3. The World Bank, Washington DC
- Gueron JM (2002) The politics of random assignment: implementing studies and impacting Policy. In: Mosteller F, Boruch R (eds) *Evidence matters: randomized trials in education research*, Brookings Institution Press, Washington DC
- Heckman J, Ichimura H, Todd P (1998) Matching as an econometric evaluation estimator: evidence from evaluating a job training program. *Review of Economic Studies* 64:605–654
- Heckman J, Ichimura H, Smith J, Todd P (1998) Characterizing selection bias using experimental data. *Econometrica* 66:1017–1099
- Howell EM, Yemane A (2006) An assessment of evaluation designs: case studies of 12 large federal evaluations. *American Journal of Evaluation* 27:219–236
- Prennushi G, Rubio G, Subbarao K (2002) Monitoring and evaluation. In: Klugman J (ed) *A sourcebook for poverty reduction strategies*, World Bank, Washington DC, pp 105–130
- Ravallion M (2001) The mystery of the vanishing benefits: an introduction to impact evaluation. *The World Bank Economic Review* 15:115–140
- Ravallion M (2005) Evaluating anti-poverty programs. *The World Bank Economic Review* 15: 115–140
- Savedoff WD, Levine R, Birdsall N, Co-Chairs (2006) When will we ever learn? Improving lives through impact evaluation. Report of the evaluation gap working group, Center for Global Development, Washington DC
- Scheuren FJ (2006) Statistics and the millennium development goals: government statistics. Discussion at the annual joint statistical meetings, Seattle, August 6–10
- Skoufias E (2005) PROGRESA and its impacts on the welfare of rural households in Mexico. International Food Policy Research Institute research report 139. International Food Policy Research Institute, Washington DC
- Wassenich P (2006) The role of monitoring and evaluation in development programs. Paper presented at the annual joint statistical meetings, Seattle, August 6–10

Samtskhe-Javakheti Roads Rehabilitation Activity

- Bakht Z (2000) Poverty impact of rural roads and markets improvement & maintenance project of Bangladesh. Paper presented at the World Bank South Asia Poverty Monitoring and Evaluation Workshop, New Delhi, June 8–10

- Chomitz KM, Gray DA (1996) Roads, land use, and deforestation: a spatial model applied to Belize. *The World Bank Economic Review* 10:487–512
- Cook CC, Duncan T, Jitsuchon S, Sharma A, Guobao W (2005) Assessing the impact of transport and energy infrastructure on poverty reduction. Asian Development Bank, Manila
- Escobal J, Ponce C (2002) The benefits of rural roads: enhancing income opportunities for the rural poor. GRADE working paper 20. Grupo de Analisis para el Desarrollo, Lima
- Grootaert C, with guidance from Calvo CM (2002) Socioeconomic impact assessment of rural roads: methodology and questionnaires. Roads and rural transport thematic group, World Bank, Washington DC
- Jacoby HG (2000) Access to markets and the benefits of rural roads. *The Economic Journal* 110 (July):713–737
- Hine J, Cundill M (1994) Economic assessment of road projects: do our current procedures tell us what we want to know? Paper presented at the International Workshop on Impact Evaluation and Analysis of Transportation Projects in Developing Countries, Bombay, December 13–16
- Lokshin M, Yemtsov R (2003) Evaluating the impact of infrastructure rehabilitation projects on household welfare in rural Georgia. World Bank working paper 3155. World Bank, Washington DC
- Miller HJ, Wu Y-H (2000) GIS software for measuring space-time accessibility in transportation planning and analysis. *GeoInformatica* 4:141–159
- Rosero-Bixby L (2004) Spatial access to health care in Costa Rica and its equity: A GIS-based study. *Social Science & Medicine* 58:1271–1284
- van de Walle D (2002) Choosing rural road investments to help reduce poverty. *World Development* 30:575–589
- van de Walle D, Cratty D (2002) Impact evaluation of a rural road rehabilitation project. World Bank, Washington DC

Agribusiness Development Activity

- Alex G, Byerlee D (2000) Monitoring and evaluation for AKIS projects: framework and options. Agricultural knowledge and information systems (AKIS) good practice note. World Bank, Washington DC
- Birkhaeuser D, Evenson RE, Feder G (1991) The economic impact of agricultural extension: a review. *Economic Development and Cultural Change* 39:607–650
- Conley TG, Udry CR (2005) Learning about a new technology: pineapple in Ghana. Paper presented at the productivity growth: causes and consequences conference of the Federal Reserve Bank of San Francisco, San Francisco, November 18–19
- Diaz JJ, Handa S (2006) An assessment of propensity score matching as a nonexperimental impact estimator: evidence from Mexico's PROGRESA program. *Journal of Human Resources* 41:319–345
- Duflo E, Kremer M (2003) Use of randomization in the evaluation of development effectiveness. Paper prepared for the World Bank operations evaluation department conference on evaluation and development effectiveness, Washington DC, July 15–16
- Evaluation Partnership Limited (2005) Impact evaluation of four CFC funded projects in Uganda: final report. United Nations Common Fund for Commodities, Amsterdam
- Feder G, Slade R (1986) The impact of agricultural extension: the training and visit system in India. *Research Observer* 1:139–161
- Friedlander D, Robins PK (1995) Evaluating program evaluations: new evidence on commonly used nonexperimental methods. *The American Economic Review* 85:923–937
- Gautam M, Anderson JR (1999) Reconsidering the evidence on the returns to T&V extension in Kenya. Operations evaluations department, World Bank, Washington DC

- Gautam M (2000) Agricultural extension: the Kenya experience: an impact evaluation. Operations evaluation department, World Bank, Washington DC
- Kumar K (1995) Measuring the performance of agricultural and rural development programs. *New directions for evaluation* 67:81–91
- Minten B (1999) Infrastructure, market access, and agricultural prices: evidence from Madagascar. International Food Policy Research Institute MSSD Discussion Paper No 26. International Food Policy Research Institute, Washington DC
- Owens T, Hoddinott J, Kinsey B (2003) The impact of agricultural extension on farm production in resettlement areas of Zimbabwe. *Economic Development and Cultural Change* 51:337–57
- Sebstad J, Snodgrass D (2004) Assessing the impact of the Kenya BDS and the horticulture development center projects in the treefruit subsector of Kenya: baseline research design. USAID, Washington DC
- Sebstad J, Snodgrass D (2005) Assessing the impact of the Kenya BDS and the horticulture development center projects in the tree fruit value chain of Kenya: baseline research report. USAID, Washington DC
- Verna S, Burnett M (1999) Addressing the attribution question in extension. Paper presented at the annual conference of the American Evaluation Association. Orlando, November 3–6
- von Oppen M, Njehia BK, Ijaimi A (1997) The impact of market access on agricultural productivity: lessons from India, Kenya and the Sudan. *Journal of International Development* 9:117–132

Georgia Regional Development Fund

- Afrane S (2002) Impact assessment of microfinance interventions in Ghana and South Africa: a synthesis of major impacts and lessons. *Journal of Microfinance* 4:37–57
- Aivazian V, Mazumdar D, Santor E (2003) Financial constraints and investment: assessing the impact of a World Bank Loan program on small and medium-sized enterprises in Sri Lanka. Bank of Canada working paper 2003-37. Bank of Canada, Ontario
- Batra G, Mahmood S (2003) Direct support to private firms: evidence on effectiveness. World Bank policy research working paper 3170. World Bank, Washington DC
- Hulme D (1997) Impact assessment methodologies for microfinance: a review. Paper prepared for the AIMS project for the virtual meeting of the CGAP working group on impact assessment methodologies, April 17 – 19
- Mosley P (1997) The use of control groups in impact assessments for microfinance. International Labour Office working paper N 19. International Labour Office, Geneva
- Rogerson CM (2004) The impact of the South African government's SMME programmes: a ten-year review (1994–2003). *Development Southern Africa* 21:765–784
- Tan H, Acevedo GL (2004) Evaluating SME training programs: some lessons from Mexico's CIMO program. World Bank, Washington DC

Some Recent Propensity Scoring Method Literature¹⁴

- Anstrom KJ, Tsiatis AA (2001) Utilizing propensity scores to estimate causal treatment effects with censored time-lagged data. *Biometrics* 57:1207–1218

¹³ ¹⁴For compiling this bibliography we conducted a wider search for citations. The numbers found were astonishing—in the hundreds. We decided not to include all of those citations here. Arguably, when the reader gets to that point there will be many more citations to explore.

- Baker SG, Lindeman KS (2001) Rethinking historical controls. *Biostatistics (Oxford)* 2:383–396
- Bakken DG, Terhanian G, Bremer J (2001) Projecting Internet survey results to the general population: a propensity score adjustment method. Paper presented at the ART forum, Amelia Island
- Cochran WG (1968) The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics* 24:205–213
- Díaz-Tena N, Potter F (2003) Nonresponse adjustments for a survey of children with disabilities using information of a responsible adult. *ASA proceedings of the joint statistical meetings* 1217–1222
- D'Agostino Jr RB, Rubin DB (2000) Estimating and using propensity scores with partially missing data. *Journal of the American Statistical Association* 95:749–759
- D'Agostino Jr RB, Craven T, Slone S, Lang W, Morgan T (2000) Estimating and using propensity scores with nonignorable missing covariates. *ASA proceedings of the epidemiology section* 85–90
- Duncan KB, Stasny EA (2001) Using propensity scores to control coverage bias in telephone surveys. *Survey Methodology* 27:121–130
- Elliott MR, Davis WW (2005) Obtaining cancer risk factor prevalence estimates in small areas: Combining data from two surveys. *Journal of the Royal Statistical Society, Series C: Applied Statistics* 54:595–609
- Ellis BH, Bannister WM, Cox JK, Fowler BM, Shannon ED, Drachman D, Adams RW, Giordano LA (2003) Utilization of the propensity score method: an exploratory comparison of proxy-completed to self-completed responses in the Medicare Health Outcomes Survey. *Health and Quality of Life Outcomes* 1:47
- Hong G, Raudenbush SW (2003) Causal inference for multi-level observational data with application to kindergarten retention study. *ASA proceedings of the joint statistical meetings* 1849–1856
- Joffe MM, Rosenbaum PR (1999) Propensity scores. *American Journal of Epidemiology* 150:327–333
- Kondratovich M (2002) Matched receiver operating characteristic (ROC) analysis and propensity scores. *ASA proceedings of the joint statistical meetings* 1904–1908
- Larsen MD (1999) An analysis of survey data on smoking using propensity scores. *Sankhyā, Series B* 61:91–105
- Marcus SM, Gibbons RD (2001) Estimating the efficacy of receiving treatment in randomized clinical trials with noncompliance. *Health Services & Outcomes Research Methodology*, 2: 247–258
- Patterson BH, Dayton CM, Graubard BI (2002) Latent class analysis of complex sample survey data: application to dietary data. *Journal of the American Statistical Association* 97:721–729
- Ridgeway G (2006) Assessing the effect of race bias in post-traffic stop outcomes using propensity scores. *Journal of Quantitative Criminology* 22:1–29
- Rosenbaum PR (1995) *Observational studies*. Springer, New York
- Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70:41–55
- Rosenbaum PR, Rubin DB (1984) Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association* 79:516–524
- Rubin DB (1985) The use of propensity scores in applied Bayesian inference. In: Bernardo JM, DeGroot MH, Lindley DV, Smith AFM (eds) *Bayesian statistics 2*, Elsevier/North-Holland, New York and Amsterdam, 463–472
- Rubin DB (1997) Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine* 127:757–763
- Rubin DB (2001) Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services & Outcomes Research Methodology* 2:169–188
- Rubin DB, Thomas N (1996) Matching using estimated propensity scores: relating theory to practice. *Biometrics* 52:249–264
- Schonlau M, van Soest A, Kapteyn A, Couper MP, Winter J (2004) Adjusting for selection bias in web surveys using propensity scores: The case of the health and retirement study. *ASA proceedings of the joint statistical meetings* 4326–4333

- Stürmer T, Schneeweiss S, Brookhart MA, Rothman KJ, Avorn J, Glynn RJ (2005) Analytic strategies to adjust confounding using exposure propensity scores and disease risk scores: nonsteroidal anti-inflammatory drugs and short-term mortality in the elderly. *American Journal of Epidemiology* 161:891–898
- Taylor H (2000) Does Internet research work? Comparing online survey results with telephone survey. *International Journal of Market Research* 42:1, 51–83
- Wassell J (2002) Causal analysis of back belts to prevent low back pain. *ASA proceedings of the joint statistical meetings* 3645–3650
- Wong EL (1999) Propensity scores in epidemiology, selectivity models in econometrics: bridging the gap. *ASA proceedings of the business and economic statistics section*, 191–196
- Wun L-M, Ezzati-Rice TM, Machlin SR, Baskin R (2003) Investigation of alternative nonresponse adjustment methods in the Medical Expenditure Panel Survey. *ASA proceedings of the joint statistical meetings*, 4612–4618
- Wun L-M, Ezzati-Rice TM, Baskin R, Greenblatt J, Zodet M, Potter F, Diaz-Tena N, Touzani M (2004) Using propensity scores to adjust weights to compensate for dwelling unit level nonresponse in the Medical Expenditure Panel Survey. *ASA Proceedings of the Joint Statistical Meetings*, 4625–4631
- Zaccaro DJ, Wolfson M, Preisser JS (2000) Use of propensity scores in a non-randomized community trial: evaluating the enforcing underage drinking laws program. *ASA proceedings of the epidemiology section*, 74–79

Chapter 13

Using Population Data Systems to Target Vulnerable Population Subgroups and Individuals: Issues and Incidents

William Seltzer and Margo Anderson

Abstract This chapter focuses on (a) presenting and discussing concepts and perspectives needed to understand how population data systems have been misused to target individuals and population subgroups, (b) presenting a review of instances, sometimes in the context of major human rights abuses, where population data systems have been used for such targeting, where such efforts were initiated, or where such targeting has been seriously contemplated or suspected, (c) a brief review of a range of safeguards that can help to prevent or reduce the impact of such misuses, and (d) in light of the material presented, a discussion of the role and responsibilities of government statistical agencies and their staffs and the statistical profession more generally.

13.1 Introduction

The purpose of this chapter is to revisit the issue of the disclosure of information about individuals and families collected by statistical agencies and programs under government auspices and its implications for the human rights of vulnerable population groups and individuals. We also assemble in one place the results of several recent studies of such disclosures, or of actions that might lead to such disclosures, placing respondents and other members of the public at added risk of death, prosecution, or the loss of liberty. In the broadest terms, the chapter has two interrelated objectives: (1) the presentation of a body of facts and (2) the presentation of a reconceptualization of a number of the issues related to disclosure and statistical confidentiality needed to understand the implications of the facts assembled. This reconceptualization is rooted in the ethical, statistical policy, and statutory origins of the idea of statistical confidentiality. In our view both parts are essential. Indeed, as we will illustrate, in the absence of this reconceptualization many of the facts themselves remained invisible for decades. Accordingly, any sound discussion of disclosure must be based on a clear understanding of the history of the concept of statistical confidentiality as it has evolved in its ethical, policy, and legal dimensions. Such a discussion must also address issues related to the concepts of harm and risk. While risk is a term frequently used in discussions of disclosure, it is most often interpreted to mean the risk of disclosure. We emphasize in this chapter two

quite different kinds of risk: risks to data providers (that is, respondents and other members of the public) occasioned by different sorts of disclosures and risks to government statistical agencies posed by such disclosures. Each of these risks entails some consideration of the harm that individuals, vulnerable population subgroups, and statistical agencies may be exposed to through such disclosures.

As the title of this chapter indicates, our focus is on one aspect of a complex set of issues. We focus on the actions of government statistical agencies, particularly the extent to which the actions of such agencies can protect vulnerable population subgroups or threaten to expose them to harm. One's understanding of the requirements of statistical confidentiality and the safeguards needed to prevent disclosure in turn defines how to protect respondents from harm. The threats we refer to relate to the personal safety of such individuals and groups from governments themselves and, in the extreme, have placed people at increased risk of major human rights abuses. Because of their potential seriousness, we believe that such risks deserve special attention, just as potentially catastrophic events in other fields are studied so that they may be avoided or otherwise protected against.

Our focus on the more serious, but hopefully rare, cases of possible disclosures related to a government's own misuse of the statistical system does not imply that we think disclosures arising from other sources should be ignored. In fact, many past studies of disclosure have concentrated on instances of someone "finding themselves" or "finding someone else," in a file, perhaps by using special information known about an individual (for example, efforts to find out about the financial condition of a former spouse) or concatenating search criteria. Such cases of disclosure or disclosure risk are important to study and protect against both from the perspective of the reputation of a statistical agency and the privacy of the affected individuals. These disclosures, however, may be likened to "mom and pop" retail efforts. By contrast, the present chapter addresses efforts that may be likened to state-sponsored wholesale enterprises.

It is sometimes suggested that, given the scope of potential private sector threats to privacy, activities of the state, particularly those involving the misuse of data gathered or held by statistical agencies, no longer pose any special threat to the individual. Such suggestions ignore the special status accorded various outputs of many federal statistical agencies by federal, state, and local courts and administrative entities in the United States.¹ Under long-standing rules of evidence, courts and administrators may take "judicial note" of the results obtained by statistical agencies, such as the Census Bureau (see, for example, *United States v. Esquivel*, 88 F.3d 722 (9th Cir.1996) in which the court stated that "census documents meet the requirements of Rule 201(b), Fed. R. Evid., in that they are 'not subject to reasonable dispute' because they are 'capable of accurate and ready determination by resort to sources whose accuracy cannot reasonably be questioned'.") "In short, courts and government officials can and have taken decisions relating to life, liberty, rights, and property relying on the official status of the information gathered by federal statistical agencies."

¹ Analogous arrangements seem to be in place in a number of other countries.

In discussing issues related to data disclosure, targeting and the resulting harms, three different types of data are usefully distinguished: *macro* data, *meso* data, and *micro* data. Since the terms are used frequently in this chapter, we will define them at the outset. *Macro* data refer to tabulated aggregates for national or large geographic areas. *Meso* data refer to tabulated data for sufficiently small geographic areas that the results can be used operationally to identify and target a vulnerable population subgroup. *Micro* data identify records for each individual. While the relative protection offered by the statistical aggregates of *macro* data and the relative vulnerability of individual records that constitute *micro* data have long been recognized, the special risks posed by *meso* data have only been explicitly examined in the past few years. Accordingly, a few more words about the nature and role of *meso* data may be useful.

As used here, *meso* data are statistical results presented at such a fine level of geographic disaggregation, whether in tabular or graphic form, that the results may be used in conducting field operations at the local level. Thus the borderline between *macro* data and *meso* data will depend in part on the size of the geographic units, the distribution of the target population among these units, and the intended operational uses. For example, census aggregates showing the number of persons in a target population for an individual small village maybe operationally useful, while similar data for a large city would need to be further broken down by tract, ward, or even block to be operationally useful.

Our analysis explores the risks associated with disclosures based on population data. We note that statistical confidentiality also protects information that statistical agencies collect on business enterprises, but given the chapter's focus, we ignore other types of disclosures and the harms that may arise from them. The federal government has also used information on individual firms and businesses obtained through statistical data collection programs carried out under a pledge of confidentiality for regulatory or prosecutorial purposes (Anderson and Seltzer 2005).

Finally, we would emphasize, as we have previously, that while addressing instances of the misuse or the diversion of government statistical systems, we do so as strong supporters of these systems, their outputs, and their staffs (see, for example, Seltzer and Anderson, 2000, 2001). Government statistical systems, and the United States federal statistical system is a prime example, gather a broad range of valuable data for policy, administration, and research. We take that as a given. We believe, however, that only by also documenting and studying things that went wrong, or nearly went wrong, are we in a position to avoid, minimize, or control for such mistakes in the future. This is a proposition that underlies all studies of error by data users and producers. The present study simply deals with a different dimension of error.

We also recognize that efforts to minimize risks to respondents may, in some cases, lead to data users perceiving a reduction in the usefulness of the available data. Some such losses to utility may be the price that data users will have to pay to ensure a continuing flow of reliable information from data providers. At the same time, there are a number of substantive, technical, and operational safeguards that data producers can use that will help in the release of more "safe" data at reduced risks for respondents.

After providing an overview of the history of the concept of statistical confidentiality (Section 13.2), we discuss the related issues of disclosure, harm, and risk that have emerged from the research on the use of government statistical agencies or programs to assist in the non-statistical task of targeting individuals or vulnerable population subgroups (Section 13.3). Section 13.4 of the chapter sets out the available evidence concerning such governmental efforts, some of which have led to serious human rights abuses. The section draws on previously reported research supplemented by more recent research by ourselves and others, some of it ongoing. It is divided into two subsections, the first dealing with examples from the United States (Section 13.4.1) and the second with examples from other parts of the world (Section 13.4.2). In Section 13.5 we describe a number of barriers to the study of disclosures, harms, and risks associated with the activities of government statistical programs. We conclude by offering a few recommendations that appear to us to flow from the material presented in the chapter (Section 13.6).

13.2 The Concept of Statistical Confidentiality: An Overview

The concept of the confidentiality of information about individuals obtained in the course of statistical investigations has foundations in ethics, statistical policy, and law. As with many aspects of official statistics, the modern concept, while over a century old, has evolved over time. This evolution has proceeded somewhat differently in terms of ethics, statistical policy, and law. Initially, these three strands evolved independently of one another. More recently they have interacted, but each retains its distinctive character. We shall address each in turn in so far as possible, given the interactions. With respect to statutory protections, we shall confine ourselves exclusively to developments in the United States.

13.2.1 Ethics

The normative foundations of statistical confidentiality have roots deep in at least two ethical principles. First is the moral imperative that we must be alert to our capacity to cause harm in our eagerness to do good. Second is the caution that personal information provided in various professional contexts should not be shared with others. The rationale for the first is the recognition that in an aura of beneficence or zeal we are sometimes blind to the harm that our actions may cause. The rationale for the second is twofold: (1) the recognition that unless those we must interact with trust us, we cannot do our job, and (2) that in doing our job we are very likely to obtain information that can be used against those we must interact with. Well, before statisticians considered such issues, those in the clergy, law, and medicine addressed them. For example, both these norms are reflected in the Hippocratic Oath that dates

back to about 400 BCE.² Of course, ethical issues as they present themselves in official statistics are also rooted in developments in the field itself as well as in developments related to experimentation and research involving human subjects.

One major development in this process was the so-called “Nuremberg code,” formulated by two expert witnesses during the course of one of the trials conducted at the conclusion of World War II. The trial dealt with some of those involved in carrying out the so-called medical experiments on concentration camp inmates.³ This Code, which was accorded the force of international humanitarian law by virtue of its adoption by the Court, stressed the absolute duty of those carrying out experiments (and research) on human subjects to avoid serious harm and to obtain voluntary informed consent from their research subjects. This Code, along with various revisions and extensions, also served as the basis for a number of normative, policy, and statutory efforts in the United States and abroad, to guide and regulate experiments and research involving human subjects during the second half of the twentieth century.

Nevertheless, initial efforts to develop a set of normative guidelines for statistics undertaken by the American Statistical Association (ASA) at the national level in the late 1940s and early 1950s concentrated primarily on issues of statistical standards, rather than non-technical matters, and confidentiality did not feature as one of the topics addressed (Clausen et al. 1954). The only reference to the issue seems to have been in a 1952 published comment listing among the provisions that might be included in a normative code for statisticians:

3. If source material is furnished him on condition that the respondent should not be specifically identified, he preserves this anonymity (Freeman 1952, p. 20).

The process by which the ASA reached a normative statement was very slow indeed, see Gibbons (1973) and Ellenberg (1983) for a review of the process. However, when it was finally approved and published the initial ASA Ethical Guidelines for Statistical Practice contained a clear and detailed statement concerning statistical confidentiality (Ireland et al., 1983, p. 5).

The Ethics Declaration adopted by the International Statistical Institute (ISI) in 1985 also contains extensive guidance on issues related to obligations to human research subjects (International Statistical Institute 1986, Part 2, Section 4), particularly Section 4.4, “Protecting the interests of subjects,” which states

Neither consent from subjects nor the legal requirement to participate absolves the statistician from an obligation to protect the subject as far as possible against potentially harmful

² As expressed by Hippocrates, “abstain (with respect to treatment) from whatever is deleterious and mischievous” and “whatever, in connection with my professional practice ... I see or hear, in the life of men, which ought not to be spoken of abroad, I will not divulge, as reckoning that all such should be kept secret” (<http://classics.mit.edu/Hippocrates/hippooath.html>, accessed on September 16, 2003). While some other translations portray Hippocrates as using more direct language, they do not alter the substance of these two principles.

³ Indeed, the War Crimes Tribunal that tried the “Doctors’ Case” at Nuremberg moved beyond medical experimentation to include a non-medical crime involving data collection, admittedly of an extreme form (Seltzer 2001, pp. 3–4).

effects of participating. The statistician should try to minimize disturbance both to subjects themselves and to the subjects' relationships with their environment.⁴

Section 4.5, "Maintaining confidentiality of records," which states

Statistical data are unconcerned with individual identities. They are collected to answer questions such as 'how many?' or 'what proportion?' not 'who?' The identities and records of co-operating (or non-cooperating) subjects should therefore be kept confidential, whether or not confidentiality has been explicitly pledged.

and Section 4.6, "Inhibiting disclosure of identities," which states

Statisticians should take appropriate measures to prevent their data from being published or otherwise released in a form that would allow any subject's identity to be disclosed or inferred.

Building on these earlier efforts, and informed by the research on the misuse of national population data systems to assist in some major human rights abuses, the most recent version of the ASA's ethical guidelines for statistical practice, approved in 1999, addressed the subject in several ways (American Statistical Association 1999). Thus, in a section of the preamble dealing with "shared values," the guidelines state

5. Adherence to all applicable laws and regulations, as well as applicable international covenants, while also seeking to change any of those that are ethically inappropriate.
6. Preservation of data archives in a manner consistent with responsible protection of the safety and confidentiality of any human beings and organizations involved (1999, p. 4).

Within the main body of the ASA guidelines, Section D, "Responsibilities to Research Subjects (including census or survey respondents and persons and organizations supplying data from administrative records, as well as subjects of physically or psychologically invasive research)," contains eight provisions. Among them, three seem to be particularly relevant to the issues of confidentiality, disclosure, and harm associated with government statistical programs:

1. Know about and adhere to appropriate rules for the protection of human subjects, including particularly vulnerable or other special populations who may be subject to special risks or who may not be fully able to protect their own interests. Assure adequate planning to support the practical value of the research, the validity of expected results, the ability to provide the protection promised, and consideration of all other ethical issues involved. Some pertinent guidance is provided . . . at the end of this document for U.S. law, the U.N. Statistical Commission, and the International Statistical Institute. Laws of other countries and their subdivisions and ethical principles of other professional organizations may provide other guidance.
4. Protect the privacy and confidentiality of research subjects and data concerning them, whether obtained directly from the subjects, from other persons, or from administrative records. Anticipate secondary and indirect uses of the data when obtaining approvals from research subjects; obtain approvals appropriate for peer review and for independent replication of analyses.
5. Be aware of legal limitations on privacy and confidentiality assurances. Do not, for example, imply protection of privacy and confidentiality from legal processes of discovery unless explicitly authorized to do so (1999, pp. 7–8).

⁴ As used here, the term in environment was referring to the political, social, and cultural environment rather than to the physical environment.

13.2.2 *Statistical Policy*

The idea of the confidentiality of individual-level statistical information first arose as a matter of statistical policy in the United States in the latter part of the nineteenth century in connection with the population census. Initially, population census returns were posted in public places for the local citizenry to review to make sure they were complete and accurate. By the middle of the nineteenth century, Congress and census officials began to be concerned about such public posting, and the practice was discontinued. Instead, census officials concerned with the integrity of the enumeration became increasingly worried about an enumerator revealing information improperly, or making use of census information for private benefit. As a result, those responsible for conducting the 1850 Census were cautioned against such abuses by the Secretary of the Interior (Bohme and Pemberton 1991, p. 3).

By the time of the Ninth Census in 1870, Census Superintendent Francis A. Walker instructed the assistant marshals responsible for the enumeration that

No graver offense can be committed by assistant marshals than to divulge information acquired in the discharge of their duty. All (information obtained) should be treated as strictly confidential ... The (Department of Interior) is determined to protect the citizen in all his rights in the present census. (quoted in Bohme and Pemberton 1991, p. 4)

The value of offering respondents the assurance that their responses in statistical inquiries would not be used against them by the Government seems to have first been articulated in explicit terms by the first Commissioner of Labor (subsequently the Commissioner of Labor Statistics) Carroll Wright. From the outset, he seemed to be cognizant of the need to maintain the confidentiality of responses from individuals and firms. At the time, Wright's investigations into labor conditions were exceedingly controversial. He relied on voluntary responses. He knew he would not receive cooperation from respondents, or confidence in his analyses, without objective and complete data. He made it a point to reassure respondents, for example, telegraphing a San Francisco businessman in 1898 that "I pledge my word as a government officer that names of your plants and of city and State in which located shall be concealed. This will be done for all plants. If senator or representative should ask for these names, he should not have them" (quoted in Goldberg and Moye 1985, p. 13).

Along the same lines, Herman Byer, Assistant Commissioner of Labor Statistics in the late 1940s, described how a subsequent Commissioner of Labor Statistics, Ethelbert Stewart, was said to have responded to congressional pressure in the 1920s to reveal identifiable *micro* data. According to Byers, Stewart was asked at a Congressional hearing to reveal the data on individual automobile manufacturers to Congress, and he refused on grounds of confidentiality. When the committee chair threatened Stewart, "Mr. Stewart, our committee will subpoena those records," Stewart responded, "You do, and I'll burn them first" (quoted in Duncan and Shelton 1978, p. 168). Like Walker, both Wright and Stewart grounded their insistence on statistical confidentiality on the requirements of statistical policy, rather than on any specific statute.

The same policy was clearly reflected in the census proclamation issued by President Taft in 1910 in connection with the Thirteenth Census:

The census has nothing to do with taxation, with army or jury service...or with the enforcement of any national, State, or local law or ordinance, nor can any person be harmed in any way by furnishing the information required (Barabba 1975, p. 27; quoted in full in Bohme and Pemberton 1991, p. 6).

With few notable exceptions, this view has also been reflected in statistical policy in the United States since then and in virtually all open societies since the end of World War II. For example, writing in 1967 about the US federal statistical system, Edgar Dunn Jr was emphatic:

legal and procedural protections against revealing information about individuals have been a very basic part of the operation of the Federal statistical programs for many, many years! ... The protection of personal privacy has long been an obsession with the directors of federal statistical programs ... (Dunn 1967, p. 23).

Although in retrospect, he overstated the situation somewhat, his statement certainly conveyed the prevailing ethos.

The most notable exceptions in the United States to Dunn's optimistic assessment were (1) the successful effort in May 1917 by Census Director Samuel Rogers to gain permission to use identifiable *micro* data from the 1910 census to assist in the investigation and prosecution those suspected of violating World War I draft registration laws (see Section 13.4.1); (2) the decision in 1939 by President Roosevelt to permit the FBI and the Army and Navy intelligence services to have access to identifiable census *micro* data on an individual, case-by-case basis (see Seltzer and Anderson (2002, pp. 39–40)); (3) the activities of Census Director J.C. Capt in 1941 and 1942 to make identifiable *micro* data in the hands of the Census Bureau available to assist in the war effort (see Section 13.3); and (4) the view held shortly after September 11, 2001 by at least some in the National Center for Education Statistics (NCES) and the Department of Education that identifiable NCES *micro* data provided an effective and appropriate means to locate or track suspected terrorists (see Seltzer and Anderson 2002).

Despite these detours, current statements of federal and international statistical policy with respect to statistical confidentiality remain unaltered. Thus the Federal Statistical Confidentiality Order issued by the Office of Management and Budget in 1997, the Fundamental Principles of Official Statistics adopted by the United Nations Statistical Commission in 1994, and the Committee on National Statistics in its 2001 publication *Principles and Practices for a Federal Statistical Agency* express a similar policy view to the effect that statistical confidentiality means that the information so obtained must be used for exclusively statistical purposes. The longer OMB and CNSTAT texts also spell out that such information should not be used by the government against those to whom the information pertains. In the words of the OMB order,

Consistent government policy protecting the privacy and confidentiality interests of persons who provide information for Federal statistical programs serves both the interests of the public and the needs of the government and society. The integrity and credibility of confidentiality pledges provides assurance to the public that information about persons or provided by persons for exclusively statistical purposes will be held in confidence and will not be used against them in any government action. Public confidence and willingness to cooperate in statistical programs substantially affects both the accuracy and completeness of

statistical information and the efficiency of statistical programs. Fair information practices and functional separation of purely statistical activities from other government activities are both essential to continued public cooperation in statistical programs (OMB 1997, p. 35047).

The relevant United Nations Statistical Commission policy states

Individual data collected by statistical agencies for statistical compilation... are to be strictly confidential and used exclusively for statistical purposes (United Nations Economic and Social Council 1994: Principle 6).

Finally, in the words of the CNSTAT report,

Data providers must be able to rely on the word of a statistical agency. An agency receives credibility with its providers by ensuring appropriate confidentiality of responses. Maintaining confidentiality, in particular, precludes the use of individually identifiable information for any administrative, regulatory, or enforcement purpose (National Research Council 2001, p. 5).

13.2.3 Statutory Protections⁵

With respect to the census, and probably to statistical operations generally, the first statutory reference to confidentiality in the United States appeared in the Tenth Census Act, adopted in 1879 and required that the enumerator to swear under penalty that "... (I) will not disclose any information contained in the schedules, lists, or statements except to my superior officers" (Holley 1938, p. 1), although as Bohme and Pemberton (1991, p. 5) point out the penalty clause of this statute pertained only to "statistics of property or business."⁶ This anomaly was corrected in the act adopted in 1889 for the next census (Bohme and Pemberton 1991, p. 6).

The acts adopted in 1899 for the Twelfth Census and in 1909 for the Thirteenth Census left the statutory concept of census confidentiality essentially unchanged (Holley 1938, pp. 2–4). That is, these census acts were designed to protect the Government from unauthorized disclosure of census *micro* data by census staff, not to safeguard individuals from the Government's use of such data to the detriment of the respondent. (Indeed, the acts adopted for the 1900 and 1910 Censuses specifically authorized the Census Director to provide to governors or courts copies of individual census returns.⁷) Nevertheless, the modern concept of census confidentiality appeared to make a major advancement in 1910 based on the language

⁵ For a parallel exposition and discussion of this subject, see Anderson and Seltzer (2007).

⁶ Lie (2002, p. 805) makes a similar point about statistical confidentiality practices in nineteenth century Norway, "it appears that it was only individual data for taxation purposes that was not to be released ... In practice, those who did not pay taxes or who had a lower social standing were more poorly protected."

⁷ According to Bohme and Pemberton (1991, p. 6), the provision of permitting the release of individual census records to governors and courts was not removed from the census legislation until 1976. However, after the 1919 Act, the provision of such information was not permitted if it was to be used to the detriment of the person to whom it related.

used by President Taft in the first presidential census proclamation quoted above in connection with the review of statistical policy. Despite the broad blanket of protection seemingly provided by President Taft's official assurance that no "person (can) be harmed in any way by furnishing the information required," as more fully discussed in Section 13.4.1, 7 years later the Census Director was advised that nothing in the Census Act itself prevented him from providing information from the 1910 Census schedules pertaining to the names and ages of suspected draft evaders to registration authorities.

The Fourteenth Census Act, adopted in 1919, marked a key advance transforming the policy visions of Francis Walker, Carroll Wright, and President Taft's census proclamation into statutory language. For the first time the census statute provided that "... in no case shall information furnished under the authority of this act be used to the detriment of the person or persons to whom such information relates ... (Holley 1938, p. 4)."

Title 13 was first enacted in 1929 as part of the Fifteenth Census Act. This act repealed all prior acts. It also carried forward the modern concept of statistical confidentiality, building on and further elaborating the statutory language of the previous act (Mitchell 1930, pp. 2–3). Section 11 of the 1929 Act stated,

That the information furnished under the provisions of this Act shall be used only for the statistical purposes for which it is supplied. No publication shall be made by the Census Office whereby the data furnished by any particular establishment or individual can be identified, nor shall the Director of the Census permit anyone other than the sworn employees of the Census Office to examine the individual reports (Public Law 13, 71st Congress, June 18, 1929).

Similarly, Section 18 of Title 13 authorized the Census Director to provide individuals with their own returns, undertake special tabulations, and

at his discretion, upon the written request of the governor of any State or Territory or of a court of record, to furnish such governor or court of record with certified copies of so much of the population or agricultural returns as may be requested. . . . *Provided, however*, that in no case shall information furnished under the authority of this Act be used to the detriment of the person or persons to whom such information relates.

In 1930, Census Director Steuart wrote to the Secretary of Commerce indicating it was his view that providing unprotected *micro* from the 1930 and subsequent census would both violate the new law and be "inadvisable as a matter of policy" and requested legal advice on the matter in light of some pending requests from other federal and local agencies (Steuart 1930). The matter was referred to the Justice Department, and the Attorney General confirmed that it was the "duty of the Director of the Census to decline" to provide such unprotected *micro* data (Mitchell 1930).

In retrospect, some of the requests seem benign, such as the request from the Woman's Bureau for individual *micro* level data relating to women's employment (Bohme and Pemberton 1991, p. 12). Other requests, also refused, dealing with lists of illiterates in several parts of the country, may well have been far less

benign.⁸ Illiteracy at that time was one of the factors considered by national and local eugenic activists as possible grounds for involuntary sterilization and similar negative eugenic efforts. Such a census-based list, as was also true of the lists of non-registrants or aliens, could be directly admitted as evidence in most legal and administrative proceedings.

In any case, by the advent of the Roosevelt administration in 1933, statistical confidentiality was embedded in the practice of major statistical agencies such the Census Bureau and the Bureau of Labor Statistics, and with respect to data collected by the Census Bureau under the provisions of Title 13, in the law as well.

Nevertheless, as War broke out in Europe in fall of 1939, the Roosevelt administration found the constraints of Title 13 burdensome and sought to amend it to support national defense. As recounted elsewhere (Seltzer and Anderson 2002, pp. 37–38), the Departments of Justice, War, and Navy attempted to amend Section 11 of Title 13 by inserting the following language:

Provided, however, that the records of the Bureau of the Census, including the individual reports, shall be available to the Federal Bureau of Investigation of the Department of Justice, Office of Naval Intelligence of the Department of the Navy, and the Intelligence Division of the Department of War in connection with violations of the laws against espionage and other matters relating to the national defense whenever, in the opinion of the Attorney General, the Secretary of War or the Secretary of the Navy, the public welfare would be served by according such access to said records (FDR Library, President's Official File: 3b–3c, Department of Commerce, Box 6, Folder: Commerce Department, 1939–1940, Census Bureau).

The Census Bureau under the leadership of Director William Lane Austin, and with strong support from the Department of Labor, was eventually able to convince the President, in the weeks before the 1940 Census was to go into the field, that the proposed legislation not be introduced (FDR Library, Presidents Official File (POF) 3b–3c: Folder: “Commerce Dept., 1939–40 Census Bureau”).

While Roosevelt was convinced that a legislative battle over Title 13 during a census and an election year was politically dangerous, there were strong signals that he was not going to let the issue die. In 1941, secure in their win from the 1940 election, New Dealers again looked at strengthening national defense and again looked to gain access to the individual-level information in census data. William Austin, Census Director who led the battle against the 1939 legislation, involuntarily retired at the end of January 1941.⁹

⁸ Although the opinion of the Attorney General stated that the micro data wanted was to assist in literacy programs (Mitchell 1930), Census Director Steuart in asking for a legal opinion on the matter did not cite any reason why the information was sought (Steuart 1930). Steuart did indicate that the information was sought by many local organizations and the Secretary of the Interior. The latter was both the senior government official responsible for federal education work and a former president of the American Medical Association. Steuart also indicated that it had been previously agreed to supply the requested micro data on illiteracy.

⁹ Austin reached 70 at the end of January and it required a Presidential Executive Order for him to remain as Director beyond January 31, 1941. Such Executive Orders were by no means exceptional. Indeed, both he and the federal statistical community had expected that he would be extended for an additional year so that he could complete the main work of the 1940 Census (Rice 1941).

Ignoring the universal recommendation from the statistical, social science, and user communities, that a person with professional knowledge be appointed as Austin's replacement, the nomination of J.C. Capt as Director was announced on April 22 and confirmed by the Senate, without debate, on May 13.¹⁰ Capt took office on May 22 and, almost immediately, on June 6, 1941, he initiated and obtained the support of Commerce Secretary Jesse Jones for legislation to eliminate the 1941 Census of Manufactures, to provide authority for periodic surveys for national defense needs, and to make individual-level census reports available for use in the "national defense program." Section 3 of S1627 provided,

That notwithstanding any other provision of law, any individual census report or any information contained therein may be used in connection with the national defense program under such rules and regulations as may be prescribed, with the approval of the President, by the Secretary of Commerce. No person shall disclose or make use of any individual census report or any information contained therein contrary to such rules and regulations; and anyone violating this provision shall be guilty of a misdemeanor and upon conviction thereof shall be fined not exceeding \$500 or be imprisoned not exceeding six months or both (Congressional Record (77th Cong., 1st Session), volume 87, pt. 6, p. 6969, August 11, 1941).

The Senate report accompanying the bill was explicit about the goals of the legislation:

The needs of the defense program are of such a character as to require full and direct information about specific individuals and business establishments. It is clearly the intent of Congress and of the administration to implement in every possible way the defense program. An essential part of this implementation must be through the proper use of statistical data to speed production and to provide the detailed knowledge needed for the planning of total defense. To continue to impose the rigid provisions of the present confidential use law of the Census Bureau on data now in the possession of the Bureau and that to be gathered and used for national defense would defeat the primary objects of the legislation here proposed. As a safeguard against the possible misuse of information submitted in individual census reports, the Secretary of Commerce is given the authority to prescribe with the approval of the President the terms and conditions for use in the defense program only of such information. A penalty is provided for the disclosure of any individual census report or any information contained therein contrary to the terms and conditions prescribed by the Secretary (77th Congress 1st session, Senate Rept 495, June 26, 1941, to accompany S1627).

The bill was passed in the Senate in August 1941 and went to the House, where it remained until the attack on Pearl Harbor brought the United States into the war.

Four days later, on December 11, 1941, Census Director Capt reminded the Commerce Secretary that the Census Bureau was still constrained by the confidentiality

¹⁰ Capt was an able administrator and very-well connected politically. As described in the words of one White House adviser to the President, "he is the man Harry Hopkins is to speak to you about, Harry brought him over (to the Census Bureau from the WPA) . . . to handle the political patronage (as a Confidential Assistant to Austin) . . . did a splendid job . . . (so that) the Senators and the Congressmen are all for him . . .", but as the same adviser adds, "(u)nfortunately, he has no professional background or standing in his profession. He has absolutely no statistical background" (Rowe 1941).

requirements of Title 13, and proposed to write the language of S1627 into an executive order to get around the ban. As Capt wrote,

(T)he Bureau of the Census has no authority at the present time to permit other governmental agencies to obtain from Census records information about individuals or business establishments that may be indispensable to the defense of the nation. Authority therefore is needed for the Bureau of the Census to make available for war purposes any record of information in the possession of the Bureau of the Census when directed to do so by the Secretary of Commerce.

In my judgment, it is necessary to have these powers vested in the Secretary of Commerce at once to make possible the flexible, efficient, and economical war-time operation of the Bureau of the Census in obtaining and making available statistics for planning and directing war efforts (NARA, RG40, General Records of the Department of Commerce, Office of the General Council, General Counsel's Subject and Index File, 1903–1946, Box 152, File 5706 -33. Memorandum from J.C. Capt to the Secretary of Commerce, December 11, 1941).

This time, it was the Justice Department, now under a new Attorney General, that objected to the proposal to void the confidentiality provisions of Title 13. The Attorney General's Office decided that there was no legal authority for such an Executive Order (NARA, RG40, General Records of the Department of Commerce, Office of the General Council, General Counsel's Subject and Index File, 1903–1946, Box 152, File 5706 -33. Memorandum to the file, E.T. Quigley, December 23, 1941).

Instead, the matter was handled legislatively and incorporated into Section 1402 of the Second War Powers Act enacted on March 27, 1942, which in part read,

That notwithstanding any other provision of law, any record, schedule, report, or return, or any information or data contained therein, now or hereafter in the possession of the Department of Commerce, or any bureau or division thereof, may be made available by the Secretary of Commerce to any branch or agency of the Government, the head of which shall have made written request therefore for use in connection with the conduct of the war. . .(U.S. Code Congressional Service, 1943, P.L. 507, 77th Congress, 2d Session (S2208)).

Although there were slight differences of wording between Capt's initial legislative effort of the previous year and Section 1402, with the former explicitly referring to "any individual report," Section 1402 effectively implemented Capt's and the administration's proposals. Section 1402 of Title XIV of the Second War Powers Act was repealed as part of the First Decontrol Act of 1947 (S931), Chapter 29, Public Law 29, passed March 31, 1947 (50 U.S.C.A. Appendix, Section 644a).¹¹

Indeed, after the end of World War II and during the remainder of the twentieth century there was a general tendency to strengthen the statutory protections gov-

¹¹ The act ended governmental control of the economy except in the case of commodities and products that were still in short supply or needed control during reconversion at home or abroad (e.g., rubber and sugar). The opening section of that law noted that "The Congress hereby declares that it is vital to a free economy and full production in the United States that all emergency controls and war powers under the Second War Powers Act be removed except in certain limited instances." The language continued by detailing the limited circumstances in which control could continue, with no further mention of Section 1402 (Seltzer and Anderson 2002, p. 44).

erning statistical confidentiality and extend them to a broader range of statistical programs. For example, such newly established agencies as the Bureau of Justice Statistics, the National Center for Health Statistics and the National Center for Education Statistics had, like the Census Bureau, strong agency-specific statistical laws prohibiting information about individuals gathered for statistical purposes to be used against them. Furthermore, even where courts found the statutory protections inadequate, Congress moved quickly to remedy the situation (Corcoran 1963).

The gradual evolution of the concept of statistical confidentiality over the last half century was suddenly interrupted in 2001. Since then, there have been two important statutory developments—the first a major backward step, the second a strong legislative reaffirmation of the principle of statistical confidentiality. As discussed in Section 13.4 and described in detail in Seltzer and Anderson (2002), in late October 2001, in response to the terrorist attacks of September 11 on the United States, the USA Patriot Act became law. Title V of that multifaceted legislation was labeled “Removing Obstacles to Investigating Terrorism” and included Section 508, “Disclosure of Information from NCES Surveys.” This Section allowed the Justice Department to obtain and use “reports, records, and information (including individually identifiable information)” relevant to an authorized investigation or prosecution of domestic or international terrorism or terrorism that transcends international boundaries in the possession of the National Center for Education Statistics, notwithstanding the strong confidentiality provisions of the 1994 National Center for Education Statistics Act. As part of a major overhaul of federal education research, the NCES Act of 1994 was effectively repealed in late 2002, but the confidentiality protections of that Act, as amended by the Patriot Act, were made part of the Education Sciences Reform Act of 2002 (107th Congress, Public Law 279, November 5, 2002), the new legislation governing education research, including NCES.¹²

Even as Congress approved Section 508 of the Patriot Act, relaxing the confidentiality protections for education statistics, virtually without debate, in 2001, and the recodification of this relaxation in November 2002, again apparently without debate, Congress also enacted in late 2002 the first government-wide law specifically protecting statistical confidentiality. This legislation, “Title V—Confidential Information Protection and Statistical Efficiency” of the E-Government Act of 2002 (107th Congress, Public Law 347, December 17, 2002), is a clear and unambiguous

¹² This was accomplished by Section 401 (a) (6) of the Education Sciences Reform Act, which transfers all the subsections (including subsections (a) through (c) (of the 1994 NCES Act)) and redesignating such subsections as subsections (c) through (e), respectively, at the end of section 183 of this Act. Although the Patriot Act went unmentioned in PL 107-279, section (c) of the 1994 NCES Act was the provision inserted by the Patriot Act. As a result, section 183 (e) of the new law contained the substantive content of section 508 of the Patriot Act. PL 107-279 also extended the confidentiality protections accorded to information collected for statistical purposes by the 1994 NCES Act to research and evaluation purposes.

statement of policy and law.¹³ Because this law is both recent and important, we describe it in some detail.

Subtitle A of Title V, “Confidential Information Protection,” consists of three sections: Section 511, which contains a statement of findings and purpose, Section 512, which spells out the limitations on use and disclosure of data and information under the act, and Section 513, which sets out the fines and penalties for the disclosure of such confidential data and information.¹⁴ Four of the provisions of Section 511 are particularly relevant to the present chapter:

The Congress finds the following:

...

(2) Pledges of confidentiality by agencies provide assurances to the public that information about individuals or organizations or provided by individuals or organizations for exclusively statistical purposes will be held in confidence and will not be used against such individuals or organizations in any agency action.

(3) Protecting the confidentiality interests of individuals or organizations who provide information under a pledge of confidentiality for Federal statistical programs serves both the interests of the public and the needs of society.

(4) Declining trust of the public in the protection of information provided under a pledge of confidentiality to the agencies adversely affects both the accuracy and completeness of statistical analyses.

(5) Ensuring that information provided under a pledge of confidentiality for statistical purposes receives protection is essential in continuing public cooperation in statistical programs.

The statements of legislative purpose set forth in Section 511 are equally relevant:

(1) To ensure that information supplied by individuals or organizations to an agency for statistical purposes under a pledge of confidentiality is used exclusively for statistical purposes.

(2) To ensure that individuals or organizations who supply information under a pledge of confidentiality to agencies for statistical purposes will neither have that information disclosed in identifiable form to anyone not authorized by his title nor have that information used for any purpose other than a statistical purpose.

(3) To safeguard the confidentiality of individually identifiable information acquired under a pledge of confidentiality for statistical purposes by controlling access to, and uses made of, such information.

¹³ There is a degree of uncertainty as to whether the provisions of the E-Government Act prevail over the provisions of the Patriot Act now embedded in the Education Sciences Reform Act. The E-Government Act became law in mid-December 2002, while the Education Sciences Reform Act became law in early November 2002, so that under the normal rules of legislative construction, PL 107-347 (the E-Government Act), as a later expression of legislative intent, would prevail over PL 107-279. Moreover, while Section 504 of Title V of PL 107-347 does discuss the effect of this law on other laws, there is no mention of the Patriot Act, PL 107-279, or any other law specifically related to the Department of Education or NCES. This would seem to imply, therefore, that Congress intended that Title V of the E-Government Act should prevail over all previous laws not specifically mentioned in Section 504, including the Patriot Act and the Education Sciences Reform Act. Nevertheless, the final decision on the issue, if it ever came to a test, would be up to first the Attorney General and then the federal courts.

¹⁴ Subtitle B of Title V permits the sharing of identifiable micro data pertaining to businesses among designated federal statistical agencies exclusively for statistical purposes.

When read in context with the definition of “non-statistical activities” that includes “any administrative, regulatory, law enforcement, adjudicatory, or other purpose that affects the rights, privileges, or benefits of a particular identifiable respondent” (Section 502 (5) (A)) or “is the subject of information requested or required to be supplied to an agency” (Section 502 (6)), the language of Section 511 sets out a policy standard that is fully consonant with current ethical norms. This standard recognizes that statistical confidentiality is not simply about disclosure, but also about how the government uses data and information collected for a statistical purpose.

The use and disclosure limitations contained in Section 512 deal separately with statistical data and information and data or information for non-statistical purposes. With regard to the former, the statutory protection, unless subsequently set aside by future legislation, is unequivocal

(a) USE OF STATISTICAL DATA OR INFORMATION—Data or information acquired by an agency under a pledge of confidentiality and for exclusively statistical purposes shall be used by officers, employees, or agents of the agency exclusively for statistical purposes.

(b) DISCLOSURE OF STATISTICAL DATA OR INFORMATION

(1) Data or information acquired by an agency under a pledge of confidentiality for exclusively statistical purposes shall not be disclosed by an agency in identifiable form, for any use other than an exclusively statistical purpose, except with the informed consent of the respondent.

(2) A disclosure pursuant to paragraph (1) is authorized only when the head of the agency approves such disclosure and the disclosure is not prohibited by any other law.

(3) This section does not restrict or diminish any confidentiality protections in law that otherwise apply to data or information acquired by an agency under a pledge of confidentiality for exclusively statistical purposes.

With respect to the use of data or information for non-statistical purposes, Section 512 (c) provides the following rule:

A statistical agency or unit shall clearly distinguish any data or information it collects for nonstatistical purposes (as authorized by law) and provide notice to the public, before the data or information is collected, that the data or information could be used for nonstatistical purposes.

Despite its strong and clear language, the effectiveness of Title V will depend in part on the rules each agency develops for its implementation. This is particularly true with respect to Section 512 (c) where the phrases “clearly distinguish” and “provide notice to the public” could be subject to a range of interpretations. Since the degree of clarity and extent of notice may be considered by some agencies to be inversely correlated with expected response rates, agency staff may find themselves torn between full compliance with the intent of Title V and more utilitarian considerations. In these circumstances, the Office of Statistical Policy within the Office of Management and Budget (OMB) has an important role to play as Title V specifically accords coordination and oversight responsibilities for the implementation of Title V to OMB.

In closing this discussion of the statutory protections accorded to statistical confidentiality, it is important to remember, and the events of the past several years underscore the point, that the book on statutory protections is always open. Laws

related to statistical confidentiality can be and have been changed. However, Title V does provide a strong base for mobilizing efforts to protect against future efforts to weaken the concept.

13.3 Understanding Disclosure, Harm, and Risk in the Context of Statistical Confidentiality

The term “disclosure” as frequently used in the context of discussions of privacy and confidentiality is a narrow one. This narrow concept focuses on the *identification* of individual data providers, by individual curious, profit-seeking, or malignant persons or entities, in sets of *micro* records maintained by statistical agencies. This narrow focus by statisticians is understandable, since many issues of identification, and disclosure so defined, can be readily addressed by a variety of statistical approaches and methods. However, such a focus frames the issue far too narrowly as the discussion of ethics, statistical policy, and US statutory protections just presented makes clear. Unfortunately, this narrow focus on the identification issue may cause us to neglect other, equally important concerns. Although identification is an important issue, two other questions are at least as critical: (1) who is attempting the identification and (2) why is the information being sought.

As noted in Section 13.2, the concept of statistical confidentiality explicitly takes into account the “who” and the “why” of disclosure as well as issues related to identification. In particular, the concept of statistical confidentiality as it has evolved places a special emphasis on protecting against disclosure to non-statistical government agencies and programs (the “who” issue) for their use to adversely affect individual “rights, privileges, or benefits” (the “why” issue). Related to the “why” issue and sometimes explicitly referred to in ethics and policy statements is the topic of harm. (Recall that President Taft used the concept in his 1910 Census proclamation.) Furthermore, as the evidence shows it is not some anonymous intruder that has been the source of the most serious efforts to misuse data collected by statistical agencies to pursue individuals and population subgroups, but the civilian and military investigative and intelligence agencies of governments.

Any discussion of the harm arising from a given action, immediately raises the issue of responsibility in the context of multiple causation. As noted elsewhere,

the determination of whether or not a specific activity gives rise to harm may itself be an ethical trap that can catch those who are unprepared. For example, one of the defenses offered by the doctors who carried out medical experiments on concentration camp inmates during WWII was that the research subjects “were doomed to die anyway” (Caplan 1992, p. 266). Indeed, no harm is often interpreted as “no added harm” or “no added risk of harm.” Along the same lines, it has been said that the Dutch Jews would have been rounded up for deportation and the Japanese Americans would have been interned, regardless of the actions of those involved with the data systems that assisted in these endeavors. Implicit in these statements is the view that the statisticians and the others involved in these data systems didn’t really cause harm and so did nothing really wrong. Using almost the same language, Jones (1981, p. 207) observed that a number of health officials associated with the Tuskegee syphilis study seemed to justify the continuation of that study, even after penicillin became

available, on the grounds that the research subjects “were now beyond medical help, thus strongly implying that no real harm had been done. . .” (Seltzer 2001).

Governmental statistical programs, given their long-term interest in promoting public confidence, are probably aware that such strained interpretations are unlikely to be productive over the long run. It may also be recalled that this “they would have died anyway” defense was rejected by the judges at the Nuremberg Trial. Moreover, the ethical principles that were developed in the course of that Trial placed a number of responsibilities on future experimenters. Among these were a duty to take into account the serious harm they can cause research subjects. Given the inherent power of governments and the special legal status accorded to the outputs of federal statistical programs, it is only prudent that examinations of disclosure risks associated with these programs also consider potential harms flowing from such disclosures.

A special set of disclosure issues arise when the target of adverse governmental action is not specifically named individuals, but all members of a vulnerable population subgroup. Historically, in the United States, most of the leadership in the federal statistical system has taken the view that the statistical system has no special obligation to protect against subgroup harm in contrast with individual harm. As a result, statistical confidentiality and disclosure issues have been addressed largely in terms of individually identifiable persons or entities (that is, *micro* data). Defenses of the Census Bureau’s role in the forced expulsion of the Japanese American population from the West Coast were often expressed in these terms (see, for example, Barabba (1980) and Dedrick (1981)).

Some realization that this narrow view, strictly interpreted, is inadequate seems to be emerging, perhaps influenced by a greater awareness of the direct operational use of census *meso* data to, for example, help target Jewish neighborhoods in Amsterdam for attack in early 1941 (Seltzer 1998, pp. 515, 525) and assist the US Army in early 1942 in locating Japanese American individuals and families living in the West Coast (see, Seltzer and Anderson (2000) and Section 13.4.1). Indeed, the concept *meso* data was introduced to identify a specific class of aggregated data that was effectively indistinguishable from identifiable *micro* data in terms of the harm which can be inflicted on those targeted.

However, even before the *meso* data concept was introduced, statisticians in the United States and elsewhere had begun to note the problem presented by the use of small-area statistical data to target vulnerable population subgroups. For example, Begeer and de Vries (1987), writing from a Dutch perspective, observed that the use of aggregate census data to target vulnerable population subgroups, while perhaps not a technical violation of traditional census confidentiality provisions, represented a highly suspect use of population census data. Similarly, former Census Director Barbara Bryant, while continuing to cast the use of the 1940 Census tabulations to target Japanese American largely in terms of the traditional, narrow view of statistical confidentiality, concluded by observing that

The Japanese-Americans who lost property and were interned simply because of their ancestry see things differently, indeed. Their devastating fate has been cited repeatedly by critics as a breach in spirit, if not in fact, of census confidentiality (Bryant and Dunn 1995, pp. 32–33).

Both Beeger and Bryant (and their co-authors) were merely reflecting the general proposition that simply following the letter of the law will not always lead to a result that is either consistent with the spirit of the law or one that can be easily reconciled with related policy statements or our ethical responsibilities.

In these circumstances, it is important that policies and practices aimed at promoting statistical confidentiality take into account the possible harms arising from *meso* data in addition to those associated with identifiable *micro* data. Accordingly, in our analysis we cover targeting involving either *meso* or identifiable *micro* data. The recent increased caution of the Census Bureau with respect to the release of small-area statistics would also seem to be a reflection of a similar concern.

13.4 The Use of Population Data Systems to Target Individuals and Population Subgroups

Table 13.1 presents, in highly summarized form, an updated listing of instances where efforts were made by national states to use a population census, a population registration system, or a related data system to target vulnerable population subgroups (or individuals) for adverse action.

We would stress that among the cases listed there is a wide range in severity of the consequences for the individuals and groups so targeted or identified. In some cases, targeting was part of a genocidal program. In other cases, the potential consequences were far less grave. Also some of the instances cited were fully implemented examples of targeting, while other present intentions that were never fully implemented. Furthermore, given the range of the time periods and countries covered, there is wide variation in the extent to which each data gathering activity listed was subject to statistical confidentiality legislation. All the cases listed do have two features in common: (1) they involve a population data system that was part of the national statistical system, or was created under the auspices of the national statistical authorities; and (2) in each case targeting was attempted or was an explicit or implicit goal. Our justification for using such a broad definition is simple. In view of the gravity of some of the examples, both for those targeted and for the statistical programs, agencies, and staffs involved, we consider that full exploration of the historical record is important so that we can assure consider that full exploration of the historical record is important so that we can assure that we have done all we can to avoid any new misuse by national or local governments.

Indeed, we view Table 13.1 as a sort of “incident report” similar in nature to ones used by those responsible for avoiding other types of public disasters in such fields as nuclear energy production, space exploration, or commercial aviation. In the study of rare events, with highly adverse consequences, it is important to examine not only the gravest disasters, but also the minor disasters, near misses, and problems avoided, to develop sound policies and practices. The scope of Table 13.1 is, however, limited by the exclusion of examples that are based solely on suspicion or suggestion, without some evidence that efforts were initiated to use a population data system for targeting or to otherwise significantly increase risks of vulnerable individuals or population subgroups to such targeting. (We should also clarify

Table 13.1 List of cases where population data systems have been used to target individuals or population subgroups, where such efforts were initiated, or where such targeting has been seriously contemplated³⁰

Place	Time period	Targeted individuals or groups	Data systems involved	Type of data ³¹	Human rights violation or presumed state intention	Source
Australia	Nineteenth & early twentieth centuries	Aborigines	Population registration	Mi	Forced migration, elements of genocide	Kraly & McQuilton 2005
China	1966–1976	Bad-class origin during cultural revolution	Population registration	Mi	Forced migration, instigated mob violence	Qin 2004
France	1940–1944	Jews	Population registration, special censuses	Ma, Mi	Forced migration, genocide	Rémond 1996, Seltzer 1998
Germany	1933–1945	Jews, Roma, and others	Numerous	Ma, Mi	Forced migration, genocide	Seltzer 1998
Hungary	1945–1946	German nationals and those reporting German mother tongue	1941 Population Census	Mi	Forced migration	Gal 1993
Netherlands	1940–1944	Jews and Roma	Population registration system	Ma, Me, Mi	Forced migration, genocide	Seltzer 1998
Norway	1845–1930	Samis and Kvens	Population censuses	Ma, Mi	Ethnic cleansing	Lie 2002
Norway	1942–1944	Jews	Special census & proposed population register	Ma, Mi	Genocide	Seltzer 1998, Sjøbye 1998
Poland	1939–1943	Jews	Primarily special censuses	Ma, Mi	Genocide	Seltzer 1998
Romania	1941–1943	Jews and Roma	1941 Population Census	Ma, mi	Forced migration, genocide	Black 2001

Rwanda	1994	Tutsi	Population registration	Ma, Mi	Genocide	des Forges 1999
South Africa	1950–1993	African and “Colored” populations	1951 Population Census and population registration	Mi	Apartheid, voter disenfranchment	McNeil 2002
United States	Nineteenth century	Native Americans	Special censuses, population registers	Ma, Mi	Forced migration	Seltzer 1999
United States	1917	Suspected draft law violators	1910 Census	Mi	Investigation & prosecution of those avoiding registration	Seltzer & Anderson 2003,2007
United States	1941–1945	Japanese Americans	1940 Census	Ma, Me, Mi (?)	Forced migration & internment	Seltzer & Anderson 2000, 2003
United States	2001–now	Suspected terrorists	NCES Surveys and ³² administrative data	Mi	Investigation & prosecution of domestic & international terrorists	Seltzer & Anderson 2002
United States	2003	Arab–Americans	2000 Census	Me	Unknown; stated purpose according to the U.S. Department of Homeland Security “to determine needs for airport signage.”	El Badry & Swanson 2007 Seltzer, 2005b
USSR	1919–1939	Minority populations	Various population censuses	Ma, Mi	Forced migration, punishment of other serious crimes	Blum 2000

³⁰ The time periods and intended targets specified refer only to those studied in the sources cited.

³¹ Ma=macro, Mi=micro, Me=meso. See text for definitions.

³² National Center for Education Statistics.

that though Table 13.1 refers to *macro* data in the body of the table, applications involving only the use of *macro* data, without some intention to also make use of meso or *micro* data, were omitted from Table 13.1.¹⁵) Also excluded from Table 13.1 are what we suspect are numerous unreported incidents in which government statistical agencies in the United States and elsewhere have quietly and successfully resisted explicit efforts to use the statistical system to target individuals and vulnerable population subgroups. It would seem useful for such statistical agencies to keep a systematic record of these thwarted efforts and to periodically report on them publicly.

At this point Table 13.1 contains 18 cases. Underscoring the rapidly evolving nature of this line of research, we note that in 2001 the first time the equivalent of this table was compiled, ten incidents were listed (Seltzer and Anderson 2001, p. 487). The additional cases now included relate to (a) the Australian Aborigines, (b) the population registration system in China during the Cultural Revolution, (c) the 1941 Hungarian Census, (d) Norwegian population censuses in the nineteenth and early twentieth centuries, (e) the South African 1951 population census and population registration system, (f) the 1910 US population census, (g) the recent effort made to use information collected by the U.S. National Center for Education Statistics under a pledge of statistical confidentiality to investigate and prosecute terrorism, and (h) the use by the U.S. Department of Homeland Security of *meso* data from the 2000 US population census in an apparent effort to target Arab Americans by detailed ancestry.

As is clear from Table 13.1, efforts to misuse population data systems to target vulnerable population subgroups, along with actual misuse have occurred in both totalitarian and democratic countries, although in democratic societies such misuses tended to occur primarily in times of national stress. In addition, the ensuing human rights abuses tended to be milder in democratic than totalitarian states. Moreover, the fact that two western democracies, the United States and Norway, are associated with five and two, respectively, of the listed incidents, is certainly more a reflection of the research efforts by American and Norwegian scholars (and the comparative openness of their national and statistical archives) than any differences across countries in the degree to which confidentiality pledges are honored or abridged.

Population registration systems were involved in 8 of the 18 cases listed in Table 13.1, regular decennial censuses in 8 cases, special censuses in 4 cases, and other or unspecified systems were involved in 2 cases. (The numbers total to more than 18 because in several of the cases listed, more than one data system was used in the targeting.)

Although the possibility of population census-based targeting frequently receives much attention in the press and is the cause of much public fear, the record seems

¹⁵ Although we exclude misuse or potential misuse attributable solely to *macro* data from the scope of this chapter, it should be understood that several European countries have legal or regulatory barriers that prevent government programs, including statistical programs, from compiling or storing even *macro* data using definitions or classifications that permit the identification of potentially vulnerable population subgroups. Many of these laws and regulations seem to have been put in place as a reaction to the events of World War II.

to be clear that population registers are an equal if not greater potential threat. Population registers were associated with such well-known gross abuses of human rights as the Jewish Holocaust in the Netherlands (over 70 percent of the resident Dutch Jews killed), Apartheid in South Africa, the Cultural Revolution in China, and the 1994 Rwandan genocide.

The targeted groups in the 18 episodes listed in Table 13.1 include racial and ethnic minorities (Jews, Roma, Samis, Kvens, Tutsi, and Japanese Americans), lingual minorities (German speakers in Hungary in 1945 and 1946), indigenous populations (Australian Aborigines and Native Americans), subject populations (the African and “Colored” populations in South Africa), socially defined outcasts (those from a “bad” social class in Maoist China), and legal outcasts (suspected draft registration violators in the United States in World War I and suspected terrorists in the United States after 9/11).

In terms of geographical scope, all regions of the world are represented in Table 13.1, except Latin America and Western Asia. It is quite possible that this geographic variation is simply an artifact of the limited research on the use of *meso* and *micro* data for targeting in these two regions. For more details about the individual incidents, see the individual sources cited in Table 13.1. In addition, Seltzer and Anderson (2001, 2003) and Seltzer (2005) provide some information about each of the listed events, except for the cases of the Chinese and South African population registration systems and the 1941 Hungarian Census.

The point of Table 13.1 is not to discourage the collection and use of population statistics. Rather, it is intended to remind those proposing to gather such data that they carry a heavy obligation to ensure that the systems they develop do not easily lend themselves to kinds of misuse portrayed in Table 13.1 and that continued attention be given to the prevention of misuse. Failure to respect these obligations, as discussed below, can lead to the public’s refusal to provide complete and accurate responses. This, in turn, can deprive the government and all other data users with the statistical data they need.

We briefly discuss these 18 examples individually in Sections 13.4.1 and 13.4.2. Toward the end of Section 13.4.2 we briefly discuss some suggestive and speculative cases not included in Table 13.1 as worthy of further research. We note that several of the new cases now included in Table 13.1 were among the suggestive and speculative cases identified in our 2001 assessment (Seltzer and Anderson 2001, pp. 501–503).

13.4.1 The US Record

To date, five examples have been identified in the United States where efforts were made to use or permit the use of data gathered through government statistical programs to assist in targeting individuals or population subgroups for possible adverse action, including criminal prosecution or forced migration. These are (a) the use of regular decennial population censuses and various federally mandated special censuses in connection with policies and programs directed against Native Americans in

the nineteenth century (Seltzer 1999), (b) the proactive assistance provided by the Census Bureau to the Justice Department and local authorities in identifying and prosecuting draft registration violators in 1917, (c) the actions of the Census Bureau in support of the forced migration and internment of Japanese Americans after the entry of the United States in World War II (Seltzer and Anderson 2000, 2001), (d) the use of the USA Patriot Act to permit the use of data collected by the National Center for Education Statistics under assurances of confidentiality to be used for the investigation and prosecution of terrorism (Seltzer and Anderson 2002), and (e) the apparent use by the U.S. Department of Homeland Security of *meso* data at the level of five-digit postal codes from the 2000 US population census to target Arab Americans by detailed ancestry (EI-Badry and Swanson, 2007, Habermann 2006, Seltzer 2005).¹⁶ As the accounts which follow demonstrate, there is considerable variation among these incidents in the degree to which statistical agency staff and management took part in targeting activities.

13.4.1.1 The Nineteenth Century Native American Experience

The nineteenth-century experience involving Native Americans, more fully described in Seltzer (1999) and summarized in Seltzer and Anderson (2001), did not primarily involve issues of disclosure. However, a number of treaty and administrative censuses carried out by the federal government were used for mixed statistical and administrative uses, and specifically to assist in planning the forced migration and the identification of individuals. For example, according to DeRosier (1970, p. 137) a “careful census” of the Choctaw nation “was conducted to determine the exact number of Indians who were eligible for removal.” He noted that the results, which indicated that the nation consisted of 17,963 Indians, 151 white persons, and 521 slaves, were “important because the War Department wanted to move approximately one-third of the nation in the first group” and were used in ordering supplies to be used in the forced migration of the Choctaws in the early 1830s from their lands in the southeastern United States. (This effort closely parallels the use of local area data and maps from the 1940 Census by the head of the Census Bureau’s Statistical Methods Division where “a major problem in the (Japanese American) evacuation program was that of . . . describing areas for a predesignated number of evacuees. The desired unit of evacuation was considered to be approximately 1,000 persons” (Wartime Civil Control Administration 1943, p. 137)¹⁷).

¹⁶ The accounts presented in this chapter concerning these examples have been largely drawn from the cited sources. However, in each case some additional material is presented here. This is particularly so with respect to the role of the Census Bureau in the forcible removal of Japanese Americans from the West Coast and the related efforts of Census Bureau Director J.C. Capt to set aside the confidentiality protections of Title 13.

¹⁷ Although Dedrick’s name is not cited in Bulletin 12, the double log quality control chart comparing the estimated and evacuated populations for each of the 108 exclusion orders (Wartime

The extent to which *micro* data from some special censuses were used to establish a system of controls is illustrated by a 1901 memoir describing a census carried in the San Carlos Indian Reservation in what is now Arizona (Elliot 1948, p. 98).

In 1884 a complete census had been made, the tribes being enumerated under their head chiefs and each camp of Indians of the same tribe under its head man. Brass tags of different shapes with one shape for each tribe had been provided. The band or subdivision of a tribe was designated by a letter of the alphabet, and each [mem]ber of a band had his number, stamped by the provost officer on the tag of the proper shape and given to each Indian whose name was recorded in books kept for the purpose. Each man was required to wear his tag at all times and to produce it when called upon Any failure to comply with these regulations was severely punished, and in a short time the system worked to the perfection I found it on my arrival.

Elliot also goes on to describe how these census-based metal tags were used to identify suspects in criminal identifications (Elliot 1948, pp. 100–101). (A similar uses of metal tags dating from the same period have been described by Hochschild (1998, p. 163) in connection with workers in the Belgian Congo, by Mamdani (2001, p. 12) in connection with the Hereros in German South West Africa after their release from concentration camps in 1908, and by Kraly (personal communication 2002) in connection with the aboriginal population of Australia.)

13.4.1.2 The Use of 1910 Census Micro Data in World War I Draft Registration Prosecutions

Although not yet as well documented as some of the other US examples, it is clear that despite the existence of President Taft's proclamation that seemed to offer protection to those participating in the 1910 Population Census that no harm would come from providing the requested information to the census authorities, the Census Bureau sought and obtained permission to use *micro* data from the 1910 Census to assist in the identification and prosecution of those who did not register for the draft after the United States entered World War I. This permission flowed from an opinion of the Solicitor of the Department of Commerce, issued within 3 months of the country's entry into World War I, permitting the Director of the Census at his discretion to provide local draft boards and the Department of Justice with the names and ages of suspected selective service law violators taken from the 1910 Census *micro* records (Bohme and Pemberton 1991, pp. 10–11)¹⁸.

This opinion was issued in response to a request from the Census Bureau Director made in June 1917 (Rogers 1917a) that began

Civil Control Administration, 1943, p. 149) reproduced in Bulletin 12 is a clear indication that these estimates were his work.

¹⁸ In fact, Bohme and Pemberton (1991, p. 10) indicate that the opinion involved was issued by the US Solicitor General. We take the more conservative approach of referring to this as a departmental opinion based on the archival records we have located.

I have received numerous requests from registration officials in various parts of the United States to furnish them with information from the census records, showing the ages of men who they believe have failed to register, although between the ages of 21 and 30 . . .¹⁹

The request goes on to quote from President Taft's 1910 proclamation issued in connection with the 1910 Census²⁰

The census has nothing to do with taxation, with army or jury service, with the compulsion of school attendance, with the regulation of immigration, or with the enforcement of any national, state, or local law or ordinance, nor can any person be harmed in any way by furnishing the information required.

as well as from the statement of Census Director Durand "to each head of a family:"

Read the President's proclamation on this schedule, which emphatically states that you can not be harmed in any way by giving information to the Census.

Observing that the proclamation would seem to prevent the Census Bureau from furnishing the information desired by the registration officials, Census Director Rogers nevertheless states in explicit terms his objective:

I believe that every branch of the Government, including this bureau, should assist at the present time, so far as possible, in securing a full registration. Accordingly, it is recommended that the matter be taken up with the President, with the view to having an order issued waiving the rigid rule laid down in Ex-President Taft's proclamation, and authorizing this Bureau to supply the proper officials (both registration and Federal) who are in control of the registration and prosecution of individuals who have failed to register, with data from the census schedules, which may show the ages of such individuals.

Within 4 days, the Acting Solicitor issued the requested opinion.²¹ It gave the Census Director the authority to provide names and ages to the registration authorities based on such considerations as,

It does not appear that any person will be harmed by the furnishing of the information desired and for the purpose which it is desired . . . Other provisions of the law prohibit the Director of the Census from giving certain information in regard to the business of individuals, firms, and corporations, but these provisions do not, in my opinion, apply to information in regard to names and the ages of individuals . . . There is nothing in the law or

¹⁹ The United States entered World War I on April 6, 1917, the Selective Service Act became law on May 18, 1917, all men born between 1886 and 1896 were required to register on June 5, 1917 (a special public holiday set aside for this purpose), and Census Director Rogers' request to the Secretary of Commerce was dated June 22, 1917.

²⁰ A more extended version of the quote from President Taft in support of census confidentiality is provided by Bohme and Pemberton (1991, p. 8). They note that same quote appeared in several subsequent census proclamations.

²¹ Bohme and Pemberton (1991, p. 10) give the June 26, 1917 as the date of this opinion; a file reference in NARA, RG40, General Records of the Department of Commerce, Office of the General Council, Subject and Index Files, 1903-1947, Box 157, file 3400-124 gives June 25, 1917 as the date. It is possible that the original opinion of the Commerce Department's Solicitor was confirmed the next day by the US Solicitor General, although Census Director Rogers' letter of thanks only refers to the opinion issued by the Departmental Solicitor (Rogers 1917b).

the proclamation which manifests an intent to restrict the Government, through the head of the Department, from furnishing names and ages of individuals as recorded in the Census office (Department of Commerce 1917).²²

Although Bohme and Pemberton (1991, p. 10) refer, in general terms, to the actual use of such information in individual prosecutions, we are unaware of any systematic research dealing with such cases. Bohme and Pemberton (1991, p. 11) also refer to the use of *micro* data from the 1920 Census in connection with deportation cases. Again, the matter has not been further studied to our knowledge.

13.4.1.3 The Japanese American Experience

The first results of our ongoing research into the use of federal statistical outputs and staff in the forced migration and incarceration of the Japanese American population, the majority born in the United States and US citizens, were given in Seltzer and Anderson (2000). Some additional findings, based on further archival research, were included in Seltzer and Anderson (2002). Since this research is continuing, the present chapter cannot in any way be viewed as definitive. Rather we consider it to be an interim progress report that focuses on the correction of earlier misinterpretations and the presentation of major new findings. Given this focus, much of the detail previously presented will be omitted, although some must necessarily be repeated for understanding. We present these results under four headings: (i) the role of Census Bureau management and staff, (ii) *macro* data, (iii) *meso* data, and (iv) unprotected *micro* data: For various reactions by Census Bureau staff and leadership to this ongoing research, see Habermann (2006), Jones (2005), and prewitt (2006).

The Role of Census Bureau Management and Staff

Perhaps our most serious error of misinterpretation in our original 2000 paper was our treatment of Census Director J.C. Capt. This also distorted to some degree our presentation of the activities of Dr Calvert L. Dedrick, who was reassigned from Chief Statistician in the Census Bureau's Division of Statistical Research to be the Bureau's "West Coast representative" and provided direct assistance to the military in the round-up and internment operations.

In our earlier portrayal, based on the evidence then in hand, we focused on Dedrick as the pivotal actor in the Bureau's involvement in assisting the Army in the internment program and treated Capt as playing a more passive role, at most providing enthusiastic, if naive support, to the activities of others. Based on a number of new materials since located, it is clear that Capt, as Director, provided leadership and direction to the Bureau's involvement first in defense and then in war efforts,

²² Holley (1938, p. 4) notes that one section of the 1910 Census Act that states "That no information furnished under the provision of the next proceeding section shall be used only for the statistical purposes for which it is supplied" actually refers to data provided by various kinds of manufacturing and other establishments, and not population data.

including the Bureau's assistance to the Japanese American internment program. Three examples of Director Capt's leadership and initiative in these matters may be cited. First, as we have already described in our review of the history of statutory protections accorded to statistical confidentiality in Section 13.2.3, it is clear that Director Capt was persistent and ultimately successful in his efforts to relax the confidentiality provisions of Title 13, an effort that he began within days of becoming Director.

Second, we recall that in our earlier study after recounting an exchange in the January 1942 Census Advisory Committee meeting between Dr Leon Truesdell, the Bureau's chief population statistician, and Director Capt that included the following colloquy:

Dr. Truesdell: . . . We got a request yesterday, for example, from one of the Navy officers in Los Angeles, wanting figures in more or less geographic detail for the Japanese residents in Los Angeles, and we are getting that out. . . .

. . .

Dr. Truesdell: That Los Angeles request I just referred to asked for census tracts.

. . .

The Director: We think it is pretty valuable. Those who got it thought they were pretty valuable. That is, if they knew there were 801 Japs in a community and only found 800 of them, then they have something to check up on. . . . (Census Advisory Committee, January 1942, pp. 20–21).

We commented on the apparent naiveté of Capt's remark (Seltzer and Anderson 2000, endnote 1). Capt's reference to the number 801 was far less naive than we realized as it appears that this was the Japanese American alien population for Terminal Island according to the 1940 Census based on unpublished census tract tabulations.²³ Terminal Island in Los Angeles harbor was the site of major Navy installations. It was also home to a thriving Japanese American fishing community. After the attack on Pearl Harbor, local officials and some naval officials came to perceive the fishing community as a major security threat. The Navy forcibly evacuated all the Japanese American residents of Terminal Island in late February 1942.

Finally, we note two of Capt's telegraphic instructions sent to the West Coast during initial stages of the Bureau's co-operation with the Army. The first was sent on February 26, 1942 to Col. Karl Bendetsen, GHQ Fourth Army, San Francisco, who had the responsibility for planning and executing the actual expulsion operations.

²³ On January 28, 1942, the *Los Angeles Times* reported that "slightly more than 800" of the "2200 Japanese living on Terminal Island" were aliens ("Eviction of Jap Aliens Sought," *Los Angeles Times*, January 28, 1942, p. 1A. See also *Los Angeles Times*, January 29, 1942, p. 6A). Census Tract no. 294 encompassed the portion of Terminal Island in the City of Los Angeles. The other half of the island is in the City of Long Beach. Table A-1 reported a total population of Tract 294 of 3,831 people, of these 2,253 were "Nonwhite, Other races." The white population was reported in native born and foreign born categories, but the Nonwhite was not (U.S. Census Bureau, 1942b). To our knowledge, the Census Bureau never published a breakdown of the detailed subcategories of "Nonwhite, Other races" or the Nonwhite population by nativity.

(In his work on the West Coast, Detrick reported to Bendetsen.) This message from Capt read:

Dr. Calvert L. Detrick leaving Washington 6:30 pm today . . . He has authority to act in helping you find solution of the problems discussed with you and General Gullion yesterday (NARA, RG40, Commerce, Entry 1, Box 144, File no. 67104).

Capt's second telegram was sent on March 21, 1942, addressed to Detrick, Hotel Whitcomb, San Francisco, and reads:

Reference 9th paragraph your letter of March 19, phrase "Confidential until released" is habitual label long used by the Bureau and need not be taken too literally under emergency conditions prevailing in wartime. Your instructions confirmed by wire to Colonel Bendetsen gave you authority to act for this Bureau. Any information available to you should be used in accordance with good judgement as dictated by the needs of the national war effort. I am relying on you to be prompt, practical and effective in the performance of your duties without being hampered by old Bureau habits, precedents, and practices that are not in complete accord with the urgent, rapidly shifting necessities of the times as they develop from hour to hour. Of course you understand that it is important to the Bureau that it receive official and public recognition for all its work. J. C. Capt—Director—Census (NARA, RG40, Commerce, Entry 1, Box 144, File no. 67104).

None of this detracts from Detrick's outstanding technical and organizational skills, which he employed in full measure in the support he provided to Col. Bendetsen during the planning and operational phases of the exclusion operations. But it is now clear that Detrick's work and the other aspects of the Bureau's cooperation with the Army in this matter were carried out in line with the policies of Director Capt and in response to his overall direction.

The Role of Macro Data

In our initial report on our research into the Japanese American internment in 2000, we provided some information on the rapid production and the general dissemination of *macro* data from the 1940 Census on Japanese Americans in the days immediately after Pearl Harbor (see, for example, Seltzer and Anderson (2000, pp. 5–6)). It is now clear that in addition to the public dissemination previously described, Director Capt was also providing these data directly to the Army and Naval intelligence services as well as to Sumner Wells in the State Department. (Wells and some of his staff played an important coordinating role on behalf of the President with respect to certain war-related intelligence operations.)

On December 10, 1941, for example, Capt provided the December 9 release to Harold B. Hoskins, Executive Assistant at the State Department. On December 13, Hoskins in turn requested Capt to send additional copies of these data to officials in the Office of Naval Intelligence, Army Military Intelligence and the FBI. On December 11, Henry Field, the intelligence official, who had requested the tabulations from the Census Bureau in November, transmitted them to Undersecretary of State Sumner Welles; on December 17, Capt forwarded additional

tabulations to Welles at Field's request.²⁴ (NARA, Record Group 59, Records of the Department of State, Decimal File, 811.5011/261-1/2; 811.5011/2662 PS/LIC; 740/00115 Pacific War/419)

Along the same lines, on December 18, 1941, the Census Bureau sent a telegraphic transmission of Japanese native born, foreign born, and total population figures to the Western Defense Command (WDC) in San Francisco. The tabulations listed the Japanese population in Western Defense Command by race for minor civil division in California, Oregon, and Washington, and for the total Japanese American population for Utah, Nevada, Idaho, Montana, and Arizona (NARA. Record Group 338, Western Defense Command and Fourth Army, Wartime Civil Control Administration, Civil Affairs Division, Unclassified Correspondence, Box 7, 013.3-014.32 File 013.31 Alien).

There is, of course, nothing illegal or wrong in such actions. However, the Census Bureau defined "Japanese" as racial category from the start, while the other agencies involved—the INS, the FBI, and the military—tended to see the issue, initially at least, as one relating to Japanese aliens. During January 1942 the military accelerated the switch in its view of the problem from one of dealing primarily with enemy aliens to one that defined the problem in primarily racial terms (that is, the Census concept). Certainly, a variety of factors contributed to this change in view, which ultimately shaped the entire exclusion and internment program. The question we ask is what role did the use of the "race" concept by the Bureau in defining "Japanese" contribute to this change, coupled as it was with the energetic efforts of the Census Bureau to tabulate and disseminate data on the Japanese American population?

The Role of Meso Data

It was our understanding that the evidence presented in our 2000 study, relating to the use of *meso* data from the 1940 Census to provide direct assistance to the Army operations directed against the Japanese American population on the West Coast, was generally accepted. We have recently learned that some have questioned whether such assistance in the form of *meso* data from the 1940 Census was provided by Dedrick and the Census Bureau. These doubts seem to be attributable to the fact that the Bureau has been unable to document from its own internal records some of the assertions relating to the use of *meso* data contained in Seltzer and Anderson (2000). Accordingly, we will summarize here what we consider the strongest evidence contained in our 2000 paper related to the use of *meso* data, as well as present some of the further material documenting such use.

²⁴ Our further research has provided additional support to Field's assertion that he was trying to obtain data from the 1940 Census on the Japanese American population prior to December 7, 1941 as part of an intelligence operation conducted for the White House. However, none of his efforts appeared to relate to identified *micro* data and we continue to remain highly skeptical of his later published accounts that he obtained such *micro* data.

As we noted earlier (Seltzer and Anderson 2000, p. 7), Dedrick himself confirmed the use of detailed *meso* data from the 1940 Census to assist the Western Defense Command in the process, stating that at the end of February 1942 when he was in San Francisco, the Census Bureau was asked to provide him

a detailed cross-tabulation for even the most minute areas, the smallest areas for which data were collected. In other words, enumeration districts and in some instances cities by blocks. . . . Sheets of paper from the tabulation machines were sent out to WCCA at the hotel in Market Street in San Francisco, and became the basis for the WCCA statistical activities (Dedrick 1981, pp. 172–173).

We also noted (Seltzer and Anderson 2000, p. 7),

These recollections (of Dedrick’s) appear to be confirmed by contemporary Bureau records that included in one of its listings of war related activities for 1942, the provision of “photographic and photostatic copies of block maps for 10 cities” to the Western Defense Command (US Census Bureau 1942a, Exhibit A:12).

Finally, several reports prepared by Census Bureau staff which we cited (Seltzer and Anderson 2000, pp. 13–15) included references to the Bureau providing various types of *meso* data to the Army as part of the Bureau’s defense that identifiable *micro* were not provided (see, for example, Bohme (1975), U.S. Census Bureau (1982), and Bohme and Pemberton (1991), although there is considerable variation among these sources concerning the geographical detail involved (Seltzer and Anderson 2000, Table 13.1).

Although we have not located the urban block maps, the Bureau reported it provided to the Western Defense Command in 1942 (US Census Bureau 1942a, Exhibit A:12), and which Tom Clark apparently recalls being used in planning the round-up operations (Clark 1972), we have found examples of their rural counterparts. (As discussed in Section 13.5, there may have been an active effort to suppress certain details of these operations.)

In terms of new material, we will cite two examples of the availability and use of detailed 1940 Census *meso* data in planning the round-up of Japanese Americans. Dedrick’s Statistical Section, working under Col. Bendetsen, produced four statistical bulletins, “Japanese Population and Family Heads, by Minor Civil Division: 1940,” each dated March 16, 1942, for Arizona (Bulletin 2), California (Bulletin 3), Washington (Bulletin 4), and Oregon (Bulletin 5), respectively. Each report consists of (a) a brief introductory text stating that “The Bureau of the Census . . . furnished the data for this bulletin by special request . . .”; (b) a state map with county boundaries, showing for each county the number of “Japanese” persons and family heads; and (c) a table showing the same data by county and minor civil division (Wartime Civil Control Administration 1942a–d).²⁵

That the contents of these bulletins, coupled with more detailed cartographic information than contained in the bulletins themselves, had direct operational consequences is clear from a one-page memorandum that Dedrick, as Chief Statistics

²⁵ In the 1940 Census, outside of major cities, most minor civil divisions fully met the criteria used in defining *meso* data, that is a unit small enough that it was operationally useful in targeting individuals or population sub-groups.

Division, wrote to Col. Bendetsen. This memorandum, dated March 21, 1942, provided in tabular form 1940 Census data, by minor civil division, relating to the 281 Japanese and 48 Japanese family heads, by citizenship status, and the 40 Japanese farm operators, together with a map of Bainbridge Island with minor civil division boundaries. Dedrick's memorandum, a copy of which is reproduced as Figure 13.1,

3/21/42

EXHIBIT A

MEMORANDUM FOR COLONEL KARL R. BENDETSSEN

SUBJECT: Japanese Population of Bainbridge Island

1. The 1940 Census reported 281 Japanese as living on Bainbridge Island. They were distributed by precincts as shown in the table below. This table also gives the number of aliens and citizens, the number of heads of households (equal to the number of families) by nativity of the head, and the number of Japanese farm operators.

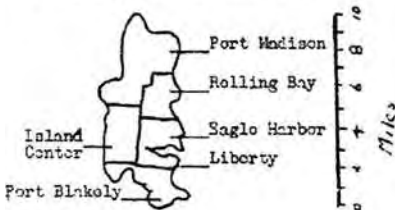
2. These figures may exceed the number of Japanese now on the island. No recent local count is available in our office for the island but figures given me for Kitsap County (including Bainbridge Island) for February, 1942 show a loss of more than one-third in the number of Japanese in that county: a reduction of from 345 in April, 1940, to 197 in February, 1942.

3. A map showing the island's precincts is given below.

	BAINBRIDGE ISLAND PRECINCTS*						
	Port Blakely	Center	Saglo Harbor	Rolling Bay	Port Madison	Liberty	All of Kitsap County
Total Japanese population...	63	23	49	19			345
Alien.....	20	8	15	6			117
Citizen...	43	15	34	13			228
Japanese heads of families...	10	7	9	3			64
Alien.....	5	4	8	3			55
Citizen...	5	-	1	-			9
Japanese farm operators...	18	8	2	10	2		49

*Liberty precinct had no Japanese in April, 1940.

Bainbridge Island Precincts



C. L. Dedrick
Chief, Statistical Division

March 21, 1942

Fig. 13.1 Memo from C. L. Dedrick to Col. Bendetsen, 3/21/1942
Source: Bancroft library, Banc MSS 67/14c FILM, Reel 12, frame 215. Reproduced with permission from the Bancroft Library, University of California, Berkeley, California. May not be further distributed or copied without written permission

constitutes one element of the Western Defense Command's "Evacuation Memorandum No. 1 (Dedrick 1942). March 21, 1942 was a Saturday. On Sunday, the Army "ordered the forcible evacuation of all Japanese from Bainbridge Island in Puget Sound . . . effective March 30" (*New York Times*, "Army Ousts Japanese Near Bremerton Yard," March 23, 1942, p. 1).

The Possible Role of Micro Data

The issue of whether identifiable *micro* data from the 1940 Census pertaining to the Japanese American population were actually used in efforts directed against them remains an open one. There seems no doubt at this time, however, that in early 1942 such a use was intended by Census Director Capt and the Congress. In 2000, we had noted (Seltzer and Anderson 2000, p. 9) that

at the same January 1942 Census Advisory Committee Meeting, in which Director Capt and his senior staff indicated that by January 10 that they were already providing tract level data on the Japanese Americans to the military, Director Capt expressed his willingness to take the next step: "We're by law required to keep confidential information by individuals. But in the end, (i) if the defense authorities found 200 Japs missing and they wanted the names of the Japs in that area, I would give them further means of checking individuals" (Census Advisory Committee, January 1942, p. 21).

Again, Capt's observation was not an idle one: we now know that he had been working for months to change the law barring such uses of confidential census data. Indeed, he renewed his efforts several weeks after the January 1942 Advisory Committee meeting by appearing before the House Judiciary Committee and presenting an "amendment to the omnibus war-powers bill . . ." to among other things, "make this and other information now obtained by the Department of Commerce under the seal of confidence available to other war agencies." (Letter from Commerce Secretary Jones to Hatton Sumners, Chairman, Judiciary Committee, dated February 4, 1942.²⁶ Reproduced in US Congress House Committee on the Judiciary "Second War Powers Act, 1942." Report No. 1765. 77th Congress, 2nd Session, February 9, 1942.)

The Judiciary Committee quickly approved this amendment, and at least according to a *New York Times* story dated February 6, 1942, Capt's intentions in introducing the amendment were clear:

The House Committee on the Judiciary before giving its approval today to the Senate adopted Second War Powers Bill, amended the measure to take back the promise it made in 1940 to that all data obtained by the Census takers would be held strictly confidential, even from other bureaus and agencies . . . Some agencies of the government want data now as a matter of national safety. They seek some of the information obtained particularly

²⁶ Capt was well-known to Chairman Sumners. Both were from Texas and, a year earlier, Sumners had himself written to Jones, urging that Capt be appointed as Census Director. (Letter to Commerce Secretary Jones, dated February 8, 1941, cited in acknowledgment from Baxter to Sumners, February 12, 1941, NARA, RG 40, Entry 12, Subject files of Undersecretary of Commerce Wayne C. Taylor, folder "Census recommendations (for Director)")

from Japanese and others who since have become enemy aliens, especially about those in costal areas from which they have been ordered evacuated by the Department of Justice . . . (Such) data, now a secret under law, government officers believe, would be of material aid in mopping up those who had eluded the general evacuation orders (*New York Times*, "Spy Data Sought From 1940 Census," February 7, 1941, p. 9)

It is also clear that the motivation cited in the *New York Times* story was not simply journalistic speculation. Some weeks later, Senator Warren Austin (R-Vermont), who was a member of the House-Senate conference committee dealing with this legislation, remarked in the course of a closed March 13, 1942 hearing of the Senate Committee on Military Affairs:

we are now considering . . . legislation that would empower the Secretary of State to requisition from the Commerce Department, Bureau of the Census, what the census shows about these people. That would give an enumeration of the Japanese and it would also give names and residences, so that, when the Army makes its evacuation it can . . . compare its list of evacuees against the census and have some knowledge of whether this has been an effective protection or not (US Congress Senate Committee on Military Affairs 1942, pp. 8–9).

Although Senator Austin misstated the administrative path by which the formerly confidential census information would be obtained under the Second War Powers Act (the law, as enacted, provided that the head of "any branch or agency of the Government" could request such information if it was needed "for use in connection with the conduct of the war," not merely the Secretary of State), the planned use, "mopping up" as the *New York Times* termed it, seems quite consistent with other available evidence. Indeed, we had previously speculated about such a selective use of *micro* data (Seltzer and Anderson 2000, p. 10).

Despite the very strong evidence that, in early 1942, both the Census Bureau and the Congress had every intention that 1940 Census identifiable *micro* data be used to assist in the expulsion and internment of the Japanese American population or similar targeting, the evidence on whether such *micro* data were actually so used is still conflicting. There is strong evidence on either side of the matter. Table 13.2 presents the evidence located through mid-2005 that seems to support or contradict actual use. We consider that the specific questions posed by Table 13.2 need further investigation.

At this point, therefore, we are not prepared to state whether identified *micro* data relating to Japanese Americans from the 1940 Census was, or was not, provided to the US Army, the Federal Bureau of Investigation, or similar agencies for war-time ". . . use against the Japanese Americans in connection with the forced migration operation on the West Coast. However, given that we now know such a use was intended by the Bureau and made lawful by congress, the importance of resolving the issue of whether or not there was actual use seems somewhat diminished, at least from an ethical perspective. Moreover, more recent research has established that in August 1943 the Census Bureau did provide identifiable microdata from the 1940 U.S. Secret Service in the form of a listing of the names, addresses, employment information, and selected demographic characteristics for all Japanese, Americans enumerated in the Washington, DC area(Seltzer and Anderson 2007)."

Table 13.2 Were micro records from the 1940 census actually used in the forced evacuation in the spring of 1942? A summary of the conflicting evidence as of mid-2005

Yes, 1940 Census *micro* data were so used

FDR's willingness in 1939 to ignore the confidentiality provisions of Title 13 on a selective basis. Legislative efforts in 1939 by the Departments of Justice, War and Navy to relax the confidentiality provisions of Title 13.

Census Director Capt's willingness, expressed in the January 1942 Census Advisory Committee meeting that "If the defense authorities ... wanted the names of (missing) Japs ... I would give them further means of checking individuals"

Capt's persistent efforts, beginning in May 1941 just after becoming Census Director and continuing through November 1941, to revive the 1939 legislative effort; his efforts in December 1941 just after Pearl Harbor to achieve the same ends by Executive Order; and finally his successful effort, in early 1942, to have an amendment inserted in the Second War Powers Bill to permit relaxing of the confidentiality provisions of Title 13.

New York Times story of February 7, 1942 about amending the Second War Powers bill that includes very suggestive language, such as, "Some agencies of the government want data now as a matter of national safety. They seek some of the information obtained particularly from Japanese and others who since have become enemy aliens, especially about those in coastal areas from which they have been ordered evacuated by the Department of Justice ... (Such) data, now a secret under law, government officers believe, would be of material aid in mopping up those who had eluded the general evacuation orders."

Senator Warren Austin's (R- Vermont) statement, during a March 13, 1942 hearing of the Senate Committee on Military Affairs, that "we are now considering ... legislation that would empower the Secretary of State to requisition from the Commerce Department, Bureau of the Census, what the census shows about these people. That would give an enumeration of the Japanese and it would also give names and residences, so that, when the Army makes its evacuation it can ... compare its list of evacuees against the census and have some knowledge of whether this has been an effective protection or not."

Capt's March 21 telegram to Dedrick in San Francisco instructing him to carry out his "duties without being hampered by old Bureau habits, precedents, and practices that are not in complete accord with the urgent, rapidly shifting necessities of the times as they develop from hour to hour." Western Defense Command (WDC) statements that Dedrick's office in San Francisco had "access to the census records of 1940" and specifically mentioning other kinds of identifiable *micro* data. Several references to special name searches in WDC documents addressed to Dedrick in the second half of 1942, for example a memorandum from Bendtsen to Dedrick, dated July 3, 1942, "Location of certain persons of Japanese ancestry" (although none of these documents specifically mentions the 1940 census).

No, 1940 Census *micro* data were not so used

No explicit evidence yet found of actual use in the various NARA record groups, Presidential libraries, and other archives so far examined.

A number of Western Defense Command documents refer to the role of Dedrick's Statistics Section as providing statistics and maps to assist in planning the evacuation and not to tracking individuals for control purposes (although there was reference to furnishing agencies concerned with resettlement with "data pertaining to the needs of individuals and families, occupations and locations").

Given the *meso* data and maps provided, the additional value of 1940 census *micro* data in the evacuation process was not clear (although the use of such *micro* data could have been used later in 1942 as part of an effort to account for "missing" Japanese Americans, as the *New York Times* story and the comments of Senator Austin indicate was the intended use of such census *micro* data).

Dedrick's strong opposition in 1939 to the proposed legislation designed to relax the confidentiality provisions of Title 13.

No record of any applications to waive the confidentiality provisions of Title 13 for individuals, as required by the 1942 law and the related Executive Order, has been found (although if the proposed use was deemed not to be statistical, it is not clear that the waiver provision would have been invoked).

13.4.1.4 Education Statistics and the USA Patriot Act

As we recounted in some detail in Seltzer and Anderson (2002), apparently in response to an idea generated internally within the National Center for Education Statistics or one of its contractors, a serious belief emerged in the Department of Education and the Justice Department in the days after 9/11 that NCES data files might provide useful leads in the investigation or prosecution of terrorism. As a result, as described in our review of the legal protection provided to statistical confidentiality in Section 13.2.3, the strong confidentiality laws that hitherto assured that NCES identifiable *micro* data could not be used to target individuals were relaxed by the Patriot Act to permit such targeting in connection with terrorism investigations and prosecutions. There is little new to add at this point beyond the possible impact of subsequent adoption of the Education Research Reform Act of 2002 and Title V of the E-Government Act of 2002, both discussed in Section 13.2.3. The fact that NCES staff appear to have been severely restricted in what they can say publicly about the matter also has hampered further understanding of how and why things evolved the way they did. In retrospect, we consider the NCES incident, a prime example of what has sometimes happened to the legal status of statistical confidentiality during times of national stress.²⁷ The fact that, as far as we know, there has been no use of NCES *micro* data for such investigative purposes may be less an indication of investigative restraint than that the most useful NCES data sets from an investigative perspective are based on samples.

²⁷ Given this lesson, we remain concerned about the continued existence of the Census Bureau's ESCAPII data files with machine-readable personal identifiers. As we previously noted (Seltzer and Anderson, 2002, pp. 45–46) as part of Census 2000 processing, “for the first time personal identifiers have been captured from the forms” (Barron 2002; Fay, 2001, p. 2). These computer files permitted the Bureau, as part of the 2000 Census evaluation program, to carry out a series of sophisticated case-by-case matches to all 280 million or so individuals enumerated in the 2000 Census (Mule, 2001).

While we understand the important role these files played in the evaluation program for the 2000 Census, as time passes the benefit-to-risk ratio associated with the further maintenance of these files, at least with 100 percent population coverage, continues to decline. Traditionally, it has been Census Bureau policy that “names and addresses are separated from the electronic files that contain an individual's answers when no longer needed, to protect the respondent's confidentiality” (U.S. Census Bureau 2001). Disposing of these files would add to the difficulty of using identifiable *micro* data from the 2000 Census to target individuals and vulnerable population subgroups. Such an operational safeguard would by now seem a prudent complement to the important but sometimes uncertain protection provided by legal safeguards.

13.4.1.5 Arab Americans and the 2000 Population Census

At the time of the US 2000 Census enumeration, the Census Bureau went to considerable trouble to reassure the responding public, particularly the various immigrant and ethnic communities, that the Japanese American experience with the 1940 Census was a special case and would never be replicated. For example, in the wake of press reports based on our study (Seltzer and Anderson 2000) of the Census Bureau's World War II involvement in the round-up and incarceration of Japanese Americans on the West Coast, Census Director Ken Prewitt (2000) wrote:

In the post-war period, important safeguards to protect against the misuse of census tabulations have been instituted, notably stronger legal provisions to protect data confidentiality and the Census Bureau's introduction of disclosure avoidance techniques. Adherence to these safeguards preclude a repeat of the 1941/42 behavior.

It was thus with considerable dismay (see, for example, "Homeland Security Given Data on Arab-Americans," *New York Times* July 29, 2004, p. A14)) that these minorities and the general public learned in 2004 that the Census Bureau had a year earlier assisted a unit of the Department of Homeland Security in obtaining detailed tabulations for each five-digit postal zip code on the number of persons reported with primary or secondary ancestry as Arab American by detailed ancestry (specifically, Egyptian, Iraqi, Jordanian, Lebanese, Moroccan, Palestinian, Syrian and two general categories, "Arab/Arabic" and "Other Arab").

It is clear from this first newspaper account of the incident and from subsequent newspaper stories over the next several months (see, for example, "Coalition Seeks Action on Shared Data on Arab-Americans" *New York Times*, August 13, 2004, p. A11; "Census Policy On Providing Sensitive Data Is Revised," *New York Times*, August 31, 2004, p. A14; and "Panel Says Census Move on Arab-Americans Recalls World War II Internments," November 10, 2004, p. A19) that there was (a) widespread and continuing interest in the Bureau's actions, (b) very great skepticism about the reasons offered by the Department of Homeland Security for seeking the zip code data on Arab Americans, (c) an immediate and continuing sense that the Census Bureau appeared to have behaved toward the Arab American population in 2003 much like its behavior toward the Japanese American population during World War II, (d) a strong sense of betrayal by the Arab American experts who had supported the Census Bureau in its work, a feeling that was shared by census experts representing other minority communities, and (e) a widespread perception that the Census Bureau's response, at least initially, was overly legalistic.

Although there was general agreement that only *meso* data were provided to the Department of Homeland Security, and that these data were publicly available on the Census Bureau's website, many ethical and policy questions were raised about the Census Bureau's actions (see, for example, El Badry and Swanson (2007) and Seltzer (2005). For a contrasting perspective, see Habermann (2005).

13.4.2 *The International Record*

13.4.2.1 Holocaust (Shoah and Porajmos)-Related Examples

Six of the fifteen examples listed in Table 13.1 relate to the targeting of Jews and Gypsies (more properly referred to as *Roma*) for segregation, forced migration, and extermination by the Nazi authorities and some of their allies during World War II. Five of these examples were discussed in detail by Seltzer (1998) and all six of them in more summary form in Seltzer and Anderson (2001). While the activities in each country are listed as one “incident” in Table 13.1, several are compound in nature. For example, while Seltzer (1998) describes the use of a 1942 special census to identify Jews in Norway as a preparatory step to their expulsion from the country, but one not carried out by the Norwegian statistical office, Søybye (1998) describes the persistent efforts of the Director-General of that office to take advantage of the situation to establish a population registration system.

We will not repeat here the detailed descriptions of these six cases contained in Black (2001), Seltzer (1998), and Søybye (1998) and the sources cited therein. As we observed in Seltzer and Anderson (2001, p. 486),

although these six cases were Nazi-inspired crimes, in only two cases, Germany itself and Poland, could the misuse of the data systems be attributed solely to Nazi initiatives. In France, Henri Buhle and Réne Camille, and in Norway, Gunnar Jahn, the heads of the statistical agencies, took advantage of the political climate of German occupation or influence, to expose vulnerable target populations to further risks by proposals to undertake major new data-gathering efforts to serve both statistical and administrative purposes (Rémond 1996, Søybye 1998).

With respect to the Netherlands we noted (2001, pp. 486, 488)

the effort at establishing a comprehensive population registration system for administrative and statistical purposes was completed even before the Nazi-occupation (Methorst 1936, Thomas 1937). In 1938 Methorst, who was then the director-general of the Dutch Central Bureau of Statistics and formerly also head of the Dutch office of population registration, reported on the rapid progress being made in the Netherlands in implementing a new comprehensive system of population registration that would follow each person “from cradle to grave” and would open up “wide perspectives for simplification of municipal administration and at the same time social research” (1938, pp. 33, 713–714). By early 1941 Methorst’s successor as head of the population registration office, J.L. Lentz, had quickly adapted this general “cradle to grave” system to create special registration systems covering the Jewish and the (Roma) populations of the Netherlands. These registration systems and the related identity cards played an important role in the apprehension of Dutch Jews and (Roma) prior to their eventual deportation to the death camps . . . The critical role of the registration system in the over-all process has been stressed by such diverse observers as the German *Generalkommissar* for administration and justice in the Netherlands in September 1941 (Presser 1969, p. 38) and the British historian Bob Moore (1997).

However, some discussion of subsequent developments related to events in France, Germany, and the Netherlands seems called for.

France

The dispute over the precise role of René Carmille, head of the French statistical service under the Pétain, referred to in Seltzer (1998, pp. 522–523), appears to continue. Led by his son, Robert Carmille, there is an ongoing effort to rehabilitate Carmille's reputation by, among other things, alleging that some of his communications to his superiors in the Vichy government were only written to deceive the government while he was deliberately using his post as head of the statistical service to sabotage the government's anti-Jewish measures. However, René Carmille's role and the role of the French statistical service itself appear to remain ambiguous at best, pending further independent research.

Germany

Our understanding of the role of the German statistical service in the Holocaust continues to benefit greatly from ongoing research by a large number of German scholars. This research, some carried out by independent scholars and some working under commission from the German statistical service, has examined operations of individual population data system and individual statisticians and demographers under the Third Reich. For some time, most of this extensive body of research was available only in German, see for example, Aly and Roth (1984) and Wietog (2001), cited in Seltzer and Anderson (2001). Fortunately, Aly and Roth is now available in English, although Wietog remains untranslated.

The Netherlands

Despite the indisputable role that the ID cards linked to the Dutch population registration system played in the Jewish and Roma Holocausts in the Netherlands and the apparent role that *meso* data from the 1930 Census played in the dot maps used to target Jewish neighborhoods, senior officials of Statistics Netherlands have persisted in the view that the statistical service, its leadership, and its outputs were uninvolved. For reasons set out in more detail previously (Seltzer 1998, pp. 523–526, Seltzer and Anderson 2001, pp. 486, 488), we think their seemingly uncategorical denials are unjustified. We would note that after the end of the war, the Dutch population registration system dropped religion and, according to Begeer (personal communication to W. Seltzer 1998), introduced a series of operational safeguards, such as decentralization, designed to make its use for targeting purposes more difficult. Despite these measures, it appears that the Dutch population had such reservations about the statistical system that they effectively forced Statistics Netherlands to abandon their plans for carrying out a population census in 1981 (Choldin 1988, p. 147, van der Laan 2000). We are also unaware of any knowledgeable, independent research on the issue that explicitly examined the degree to which the Dutch statistical service, its leadership, and its outputs in the 1930s and 1940s may have contributed to ensuing events. One issue that such research might attempt to address is the extent to

which the Dutch central statistical agency and its Director-General used the mantle of that agency's influence and prestige in the 1930s to secure the passage of the legislation needed to strengthen the county's population registration system. It may also be noted that the establishment and strengthening of national population registration systems was one of the major goals of the international eugenics movement in the 1920s and 1930s and, according to van de Kaa (1998, p. 113), Methorst was one of the "main protagonists" of the eugenics movement in the Netherlands in the inter-War period.

13.4.2.2 Other European Examples

Norway(Nineteenth Century)

A study by Lie (2002) describes the activities of the Norwegian Central Bureau of Statistics (CBS) in defining and targeting two ethnic minorities, the Samis (an indigenous people from Northern Norway, some times referred to as Lapps, a pejorative term) and the Kven (immigrants from Finland and their descendants) beginning in the later half of the nineteenth century. With respect to *macro* data, it appears that the Norwegian CBS shaped the definitions and classifications related to these two ethnic minorities to further government policies related to "Norwegianization" and boundary disputes with Finland, and related security concerns. With respect to identifiable *micro* data, Lie (2002, p. 805) reports that the CBS archives contain a number of examples during this period "where transcripts of the personal (census) forms of certain individuals and groups were made at the request of other governmental agencies . . . But the only examples found of transcripts of information from whole groups, concerns the minorities." He also cites (2002, p. 813) a specific example where a transcript was provided, on request, to the chair of the so-called "Lapp Commission," who was also a district attorney, containing individually identified personal data for "all Sami in and south of . . . Søndre Trondhejn" from the 1891 population census, including data on occupation, nationality, family ties, mental health, and ownership of property and livestock.

Hungary

In the case of Hungary, individual records from the 1941 Hungarian Census were examined at the end of World War II to target those who reported German as a mother tongue for deportation to East Germany or the Soviet Union. (It should be noted that, according to Gal (1993), those Hungarians who had actively collaborated with the Germans during World War II had already been deported or killed prior to the census-based linguistic targeting of 1945 and 1946.)

The Former USSR

At this point the best summary source that we are aware of on this experience remains (Blum (2000). As we indicated previously (Seltzer and Anderson 2001, pp. 492–493), that study "reconfirmed several instances when census *micro* data were

used to target minority population groups for forced migration and other human rights violations. It also found that the evidence seemed to indicate that by the time of the 1937 census, Stalin was relying on other data systems for *micro* data and that census was primarily used as a source of *macro* data to evaluate policies, including those of forced migration and other programs with human rights consequences.”

13.4.2.3 Elsewhere

Australia (Aborigines)

As part of an ongoing research effort aimed at documenting and better understanding the use of population data systems by national and state governments in Australia to monitor and control the Aborigines, Kraly and McQuilton (2005) have produced an initial report dealing with Victoria, first as a colony and then as a state in the early Federal period. In line with the conceptualization, used here and in our previous studies, Kraly and McQuilton present some results under the rubric of *macro*, *meso*, and *micro* data. Through the early Federal period, in Victoria at least, there seemed to be relatively little interest in *macro* data relating to the Aborigines, although unlike several other parts of the country population censuses after 1857 made attempts to cover Aborigines. By contrast, there was great interest in *micro* and *meso* data in the administration of this population, including an early emphasis on securing *micro* records with identifiable names. A population register was established to help control the population required to live in designated reserve areas and to exclude persons of mixed race from living in these reserves. The population register was also used to facilitate forced migration. Since it is known that practices varied widely across Australia during the period covered, it is important to await further research, most of it archival in nature, before attempting to draw more general conclusions.

China (the Cultural Revolution)

With respect to China, the broad outlines of the Cultural Revolution are generally known. What is less widely known is the role that the population registration system played in targeting an unknown number of victims of such human rights abuses as forced migration and mob violence, sometimes leading to death, because they were identified in the register as coming from a “bad” social class. It may be noted that one of the reforms introduced, after the excesses of the Cultural Revolution was recognized by Chinese authorities, was the elimination of social class as a variable in the population register (Qin 2004).

Rwanda (1994 Genocide)

In Seltzer and Anderson (2001, p. 493), we briefly summarized the population registration system established in the colonial period, noting that

in the 1930s this registration system was used to help fix the identity of the population in terms of the hitherto somewhat amorphous categories 'Hutu' and 'Tutsi,' primarily to assist a pro-Tutsi policy by the Belgian colonial administration based on pseudo-scientific racial grounds.

The registration system continued to function well as a statistical and administrative activity right up to the outbreak of the genocide in 1994, which made use of the system to target the Tutsi population in some areas. We also commented in connection with our discussion of needed research (2001, p. 500)

Moreover, although the use of the population register in the 1994 Rwanda genocide has been established, its role has yet to be spelled out in any detail.

Unfortunately, as far as we are aware no further research along these lines has been undertaken, although it has been clarified that the register was used in compiling lists of intended victims in the Kigali area (des Forges, personal communication to W. Seltzer 2004).

South Africa (Apartheid)

In our previous research (Seltzer and Anderson 2001, p. 501), we also identified South Africa between the 1930s and 1993 as a possible area for further research in our previous study, noting

Although de Klerk (1998, p. 74) identified the Population Registration Act as the "cornerstone" of *apartheid*, given the way that so many parts of the state administration were used in furthering systematic abuses directed at the nonwhite population and the long gestation period of the system, a full review of activities related to all population data systems, including the population registration system and regular and special censuses, seems in order.

Such a study is now underway, initiated by the work of McNeil (2002), and focused on the period 1950–1959 and, in particular, on 1950 Population Registration Act, on its early implementation, and the 1951 Census of South Africa, and on the role played by the Department of Census and Statistics, including one of its directors, J.I. Raats, in the registration system and the census in the service of *apartheid*.

In brief, in May 1948, the Nationalist party won the Parliamentary elections in South Africa, and their leader, Dr Daniel F. Malan became the Prime Minister. They were elected on platform that emphasized rigid racial separation, *apartheid*, although the term meant quite different things to different members of the government and its supporters. It appears that in one or two of the influential Nationalist documents published between 1946 and 1948, reference was made to the importance of establishing some sort of population registration system. In any event, about 6 months after the election, the Prime Minister indicated that his government might establish a system of "racial registration" (*New York Times*, "South African Hints Racial Registration," November 17, 1948, p. 7). After some delay, the government introduced legislation establishing such a population registration system in February 1950. (During the course of the Parliamentary debate the government attributed the delay to the need to send J.I. Raats, Director of the Department of Census and

Statistics, and another civil servant on a study tour to Europe and North America to collect experience on population registration in other countries.)

Among the features of this legislation of special relevance to the themes of this chapter were that (1) it provided for the establishment of a population registration system and a related system of personal identification cards, (2) each person above a minimum age was assigned to one of three main racial groups (white, “Colored”, and native) with some additional ethnic/tribal information obtained for those assigned to the latter two groups, (3) the initial assignment of “race” was to be made primarily on the basis of information to be reported in the 1951 Population Census, (4) the responsibility for assigning race to each person was given to the Director of Census and Statistics, (5) the responsibility for implementing the Act, including the registration system and the ID cards, was given to the Department of Census and Statistics, and (5) the oversight responsibility for the implementation of the Act to assigned the national Statistical Council (South Africa (Union of) 1950).²⁸

The motives for and many features of the legislation were hotly debated both in the House of Assembly and in the Senate and the final measure was not adopted until June, after the rules of debate were altered and cloture was invoked (Union of South Africa Parliament House of Assembly 1950a; Union of South Africa Parliament Senate 1950). In terms of the its initial objectives, it appears that the Act was serving a twofold purpose: first, it was an important early step in developing the infra-structure of the *apartheid* system and, second, it was specifically aimed in the short run at the “Colored” population (that is, the mixed-race and Indian populations), who the government perceived as a threat in racial and political terms. Hence, initially at least, targeting actually attributable to the 1950 Population Registration Act seemed to direct against the “Colored” population.

During the course of the debates Director Raats’ views and expertise were frequently cited by the Interior Minister and other members of the majority coalition in support of the bill and he was a major witness before the Select Committee created to finalize the bill in the middle of the legislative session (Union of South Africa Parliament House of Assembly 1950b). Indeed, on the basis of his statements at the Select Committee, it is clear that Raats, himself, was the major author of the legislation as it was introduced and that on his own initiative he designed an exemplar ID card circulated to Select Committee members.

Moreover, Director Raats and the whole South African statistical system were often invoked explicitly to provide legitimacy to the proposed registration system. For example, the Minister of Interior T. E. Dönges at the end of his lengthy opening statement introducing the bill stated:

In conclusion, I want to say that this whole system will be under the supervision of the Statistical Council. This is a council brought into being under the Statistics Act . . . I may just add that the Statistics Council has expressed itself very strongly in favor of the introduction of a population register. The fact that the Statistical Council is entrusted with the

²⁸ During the course of its passage through Parliament the legislation was amended several times by the government in response to many criticisms that arose in the debate. However, the features listed here were in the measure when it was first introduced and remained after it was adopted.

task of supervising the functioning of this scheme in general is the best guarantee of the scientific character of this system and of its bona fides . . . (South Africa. Parliament. House of Assembly 1950a, cols 2523–2534).

Despite Director Raats' role, it is clear that the information he obtained during the course of his study tour was not always correct and the legislation as drafted was technically flawed in a number of respects. Nevertheless, the legislation was passed, the 1951 Census was conducted, the Department of Census and Statistics received added posts and accommodations to support its new implementation responsibilities, and the national Statistical Council continued to approve the work being carried out. Throughout Raats' directorship, his annual reports reflect great optimism, slow progress, and the continued need for more resources. That things were far more complicated can be seen, for example, in a brief story that appeared in the *New York Times* in 1955, dealing with the implementation of the Act, which noted that "the director of the census, J.I. Raats . . . personally classified 700 persons after separate interviews" (*New York Times*, August 21, 1955, "KIMBERLEY BEGINS TO CLASSIFY RACES; Some Who Had Considered Themselves Coloreds Are Placed in Negro Bracket," p. 26). A more extended documentation of some of this classification work may be found in the report of the human rights worker, Muriel Horrell (1959).

In 1956, Director Raats left his post as director and was replaced by his deputy, Dr H.M. Stoker (South Africa. Bureau of Census and Statistics 1960). The next annual report, prepared by Director Stoker, was far more gloomy. The 1959 annual report of the national Statistical Council contains the following entry under "Population Registration:"

as from 1st October 1959, separate establishments were created in respect of the Bureau of Census and Statistics on the one hand, and the Population Register on the other. It takes note of the proposed amending legislation before Parliament in this respect and that it is the intention to terminate the supervisory function of Council regarding the Population Registration Act. It takes cognizance of the verbal information supplied by the Director, that substantial progress has been made in the work of the Population Register (South Africa. Bureau of Census and Statistics, No date, pp. 2–3).

With these changes, the added posts and office space quickly vanished and for a number of years thereafter the South African Department of Census and Statistics appears to have fallen into the shadows.

This account is only a summary of the information gathered to date on the South African experience and many gaps remain. However, even from this abbreviated and partial report, the deep and extended involvement of the South African statistical service and its leader in diverting many of the core functions of the statistical system, including the population census, to harm individuals and population subgroups is clear.

13.4.2.4 Not Yet Studied, Includes Speculative Instances

In Seltzer and Anderson (2001, pp. 501–503), we listed a number of countries and situations, outside the United States, where further research might be

productive-based indications that ranged from the suggestive to the speculative. That listing included Colonial Africa (the late nineteenth century to the 1950s), South Africa (1930s–1993), Namibia (particularly 1960s–1980s), Korea and Taiwan (1890s–1990s), countries of Eastern Europe (1948–1990s), the treatment of the *Roma* (Gypsy) population in many European countries (since 1900), persons of Japanese ancestry in several Latin American countries (1942–1945), the treatment of minority populations in China (since 1950s), the treatment of minority population groups in Finland, Sweden, and Norway, Aborigines in Australia, and suspected Irish “terrorists” in the United Kingdom, Arab “terrorists” in Israel, and Israeli “terrorists” in a number of Islamic countries. Studies dealing with three of these topics are already underway and preliminary results are reflected in Table 13.1 (the Australian Aborigines, the Samis and Kvens in Norway, and South Africa in the 1950s).

A few additional comments may be useful. First, we add to the list of areas where further research on targeting might be productive the several countries in Central and South America that were under severely repressive regimes, particularly during the 1970s and 1980s. Second, given the nineteenth century Norwegian experience with the Samis and Kvens, an examination of contemporaneous practices in Sweden with respect to these populations and other vulnerable groups seems warranted (see Rogers and Nelson 2003). Third, we reiterate our suggestion that the possibility that population data systems were used to search for suspected Irish “terrorists” in the United Kingdom, Arab “terrorists” in Israel, and Israeli “terrorists” in a number of Islamic countries be investigated. In this connection, we would suggest that practices in Northern Island particularly when it was under military control be examined. Finally, we would stress again the value of promptly undertaking studies of “targeting” by the statistical services of the countries of Eastern Europe when they were under the control of the former USSR and the Communist Party. The countries all had strong central statistical services, active programs of population data collection, and a strong potential interest in targeting. But what actually happened? Where there differences among countries? If so, what accounts for these differences? The answers to these questions may well help statistical services elsewhere cope with future threats. It appears that the time is now ripe for such research. On the one hand, state security archives have opened in a number of these countries and, on the other hand, many individuals who may have been involved with both statistical and security operations may still be alive.

13.5 Barriers to the Study of Disclosure, Harm, and Risk

This chapter focuses on substantial harms—actual or threatened—that have been identified as associated with national statistics agencies and their operations, principally through the use of *meso* and *micro* data to target individuals or members of vulnerable population subgroups. Although some of these harms have been of the gravest sort, the number of such instances identified and documented is not large in absolute terms. How much solace can be drawn from the fact that so few

cases have been so far identified? We believe, at this point, only a limited sense of security is justified.

Three broad groups of factors, in our view, account for this state of affairs. First, the subject has not been studied in depth until comparatively recently. The first study in the United States appeared only in 1998 (Seltzer 1998). Thus, the full dimensions of the problem have not been determined. So far, moreover, further study has identified further confirmatory instances.

Second, research on the topic is inherently difficult, often involving material deposited in diverse archives²⁹ and requiring multidisciplinary skills that include knowledge of governmental statistical operations, methods of historical research, and familiarity with the general history of the era involved. In addition, pertinent material may not be referenced in archival finding aids or may be discarded altogether as of no historical interest.

Moreover, the fact that the topic has only recently been identified as a distinct subject of research has in the past combined with these inherent difficulties to apparently render the issue invisible to otherwise well-qualified researchers. For example, in the course of his well-known study of the forced evacuation of Japanese Americans from the West Coast, Mortin Grodzins (1949, pp. 332–337) quotes and discusses at some length the unpublished March 13, 1942 hearing of the Senate Committee on Military Affairs without even noting that this hearing contains an explicit statement of Congressional intent concerning the use of 1940 Census *micro* data to pursue Japanese Americans missed in the initial Army round-up (US Congress, Senate, Committee on Military Affairs 1942, pp 8–9) (see Section 13.4.1 for details).

Far more recently, and even where national statistical systems and statistical confidentiality were the focus, both Duncan et al. (1993) and Seltzer (1994) underplayed or ignored the complicity of statistical systems and personnel in targeting and gross human rights abuses.

Over and above these problems of focus, effort, and expertise, other factors of a less neutral sort also operate to thwart the discovery and documentation of efforts to divert a national statistical system from its traditional mission to one of targeting individuals or members of vulnerable subpopulations. In particular, statistical agencies and their supporters, together with the involved civilian, military, and intelligence agencies, often have little interest in discussing such incidents and sometimes take active measures to suppress any knowledge about them altogether.

At times, statistical agencies and some of their supporters may be reluctant to discuss the subject over fears that any public airing of the issue will adversely impact future response rates. (In these circumstances, those who cite questionable past practices are sometimes labeled “irresponsible.”) Moreover, in many countries,

²⁹ For example, in research we have done relating the involvement of the Census Bureau in the internment of Japanese Americans and related efforts to relax the confidentiality provisions of Title 13, important materials were found not only in the National Archives holdings for the Census Bureau and the Department of Commerce, but also among the legislative and executive branch and military records maintained by the National Archives, two presidential libraries, and several academic institutions. While some material was found in multiple archives, much of it survived only in a single archive.

those non-statistical agencies most commonly associated with the use or attempted misuse of population data systems (ministries of the Interior or Justice, the military, and intelligence agencies and security police, and so on) have routine policies and practices that limit the availability and accessibility of pertinent archival material.

As difficult as it is to identify and document some cases of disclosure and misuse, it is often equally difficult to identify efforts aimed at the deliberate suppression of information about the involvement statistical agencies, their data files, and their staff in such misuse. This is particularly true when major human rights abuses take place since perpetrators frequently take wide-ranging overt actions to hide or disguise their operations.

The case of the forced migration and internment of Japanese Americans on the West Coast, discussed earlier, represents a very mixed picture when it comes to barriers to the study of disclosure. Overall, the treatment of the Japanese American after Pearl Harbor is one of the most thoroughly studied human rights abuses based, in part, on contemporary and continuing research and a very rich archival record. However, with respect to the activities and outputs of the Census Bureau there are several major gaps. For example, the archival files maintained by the National Archives and Records Administration for the Census Bureau (Record Group 29) contain relatively little information on Dedrick's operations on the West Coast and the Census Bureau's supporting activities, although numerous materials can be found elsewhere. Interestingly, among the relevant materials that Record Group 29 does contain are two notes by Dedrick, one indicating with respect to Box 219, "nothing of permanent archival value in this box. To be destroyed" (Dedrick, no date) and the other "I perceive *no need* (emphasis in the original) to keep any of the materials in Box 222" (Dedrick 1973). While both of these notes were later over-ridden by others, they do indicate a state of mind and might well help explain why almost nothing pertaining to his work as the West Coast representative for the Census Bureau between February and June 1942 were retained among Dedrick's other papers stored in Record Group 29.

A similar situation seemed to play out on the West Coast. As already described, Dedrick established and headed the Statistical Branch of the Wartime Civil Control Administration (WCCA). The WCCA was under direct control of the military and was operationally responsible for the actual expulsion of the Japanese American population from their homes and farms and maintaining them in temporary assembly centers pending their further relocation to more permanent camps. The latter camps were administered by the War Relocation Authority (WRA), an authority under civilian control. The WRA also had a statistical unit and in the words of the WRA official who at the end of the war had overall responsibility for this unit,

The WCCA maintained an elaborate statistical division which played an important part not only in planning and executing the evacuation but also in directing the subsequent movement to relocation centers. Population accounting procedures established . . . (by the WCCA) were . . . continued by . . . (the WRA). The San Francisco offices of the two organizations were on adjacent floors of the same building, and information from the Statistical Division of the WCCA was available to the WRA. There appears to have been a more or less implicit understanding by the WRA that many, if not all, of the basic evacuee records of the WCCA, particularly those relating to number, location, identification, and movement of evacuees would be transferred to the WRA . . . (Stauber 1946, p. 15).

For whatever reason, Dedrick's Statistics Division decided to withhold these materials as Stauber indicates:

In the fall of 1942, about the time the transfer of evacuees from (temporary) Assembly Centers to (more permanent) Relocation Centers was approaching completion, it became apparent, that contrary to expectations, the WCCA basic records and files would not be transferred to the WRA (Stauber 1946, p. 17).

In any case, as previously mentioned, a number of key documents related to Dedrick's and the Census Bureau's involvement in the exclusion and internment have not yet been found. These include, for example, a long letter Dedrick wrote to Capt about his work on the West Coast dated March 19, 1942. (A copy of Capt's telegraphed response was recently located in Commerce Department files (Capt 1942).) The detailed urban block level 1940 Census tabulations and block maps with information from these tabulations shown on the maps have also not been located.

13.6 Conclusions and Recommendations

As stated at the outset, this chapter has attempted to do two things. First, it presents a reconceptualization of a number of the issues related to the consideration of data access and statistical confidentiality, particularly with respect to data and information gathered by, or in the name of, government statistical programs. This reconceptualization is not an arbitrary one, rather it is derived from the detailed historical review of the ethical, statistical policy, and statutory background of the concept of statistical confidentiality presented in Section 13.2 of the chapter. The implications of this revised perspective might be summarized in these terms: any consideration of data access needs to take into account the nature of the consequences that such access might bring, particularly consequences that might threaten the life, liberty, and property of individuals and vulnerable population subgroups. Further, our responsibilities in this regard are defined not only by law, but also by a body of well-articulated statistical policy and ethics.

The chapter also presents a review of some 18 "incidents" where population data systems have been used to target individuals or population subgroups, where such efforts were initiated, or where such targeting has been seriously contemplated as well as a brief discussion of other instances of possible targeting, ranging from the probable to the speculative. Collectively, they suggest that such incidents are comparatively rare, but when they have occurred, their impact has sometimes been catastrophic. Moreover, until the topic is more broadly and thoroughly researched, any attempt to quantify the expression "comparatively rare" would be unjustified.

A number of qualitative or policy conclusions do seem warranted, which we shall express in terms of five summary recommendations. In one form or another, the first four of these recommendations have been made previously (Seltzer 1998, pp. 544–545, Seltzer and Anderson 2000, pp. 40–43, 2001, pp. 500–506), so it will be unnecessary to elaborate on them in detail. The final one is new.

13.6.1 Safeguards and Prevention Strategies

We recommend, as we have done previously, that active attention be given to developing and maintaining a robust and diverse system of safeguards against misuse together with a related prevention strategy. As we set out in more detail in Seltzer and Anderson (2001, pp. 495–500), such a system includes attention to substantive safeguards, methodological and technological safeguards, organizational and operational safeguards, legal safeguards, and ethical safeguards. Over the short run, one or another of these safeguards may weaken or fail. However, over the longer-term and collectively, they appear to minimize the harm that is likely to flow from efforts designed to use population data systems to target individuals and vulnerable population subgroups. Another way of considering prevention and safeguards strategies is to look for those factors that appear in the past to have been associated with misuse or near misuse. Table 13.3 presents such an assessment, which lists in terms of critical and aggregating factors those situations that seem to increase the possibility for respondent or group harm.

Table 13.3 Factors contributing to higher risk of population data collection effort based on potential for respondent or group harm

A. Critical factors

1. Population studied is weak or otherwise vulnerable.
2. Data gathering or research involves variables that are on “sensitive” topics, typically topics that are or can be used to identify or stigmatize one or more vulnerable groups, or use classifications that permit the identification or stigmatization of such groups.

B. Aggravating factor

1. All or substantially all of population is covered, that is, sampling is not used.
 2. Longitudinal data gathering is involved, or the activity can be linked to a longitudinal system.
 3. Participation is mandatory or is effectively coerced.
 4. Little or no input from the subject population in planning the data gathering or research activities. (The risk potential is further enhanced if there are substantial inputs in terms of expertise, staff, or funds from foreign persons or institutions.)
 5. The data gathering or research is carried out in a war, a period of civil disruption, or during or shortly after a similar emergency.
 6. Little or no attention given to organizational, operational, methodological, and technological safeguards against the misuse of information obtained for non-statistical purposes.
 7. Confidentiality assurances provided to respondents have limited or no legal basis.
 8. Ethical reviews are not carried out, are perfunctory, or are heavily influenced by utilitarian considerations.
-

Note: The presence of either or both critical factors gives rise to a presumption of risk and each additional aggravating factor present further augments of such risk. On the other hand, it should be emphasized that the presence of critical and aggravating factors does not mean that actual harm has occurred

Source: Seltzer W (2003) Data collection, Ethics Issues in. In Demeny P, McNicoll G (eds) Encyclopedia of Population. Macmillan Reference USA, New York pp 195–197

13.6.2 More Attention to Risks Associated with Meso Data

To date, the risks associated with the release of *meso* data (small area data) in tabular or cartographic form, or as *micro* records with small area identifiers have received relatively little explicit attention by those dealing with statistical policy and confidentiality issues. The problem has not been totally ignored, but is usually only addressed in the context of how the availability of small area identifiers increase the probabilities of individual identification. As we have shown, *meso* data have been used in their own right to assist in targeting operations directed by governments against vulnerable population subgroups in the past. Accordingly, we would urge those doing research on confidentiality issues and those in statistical agencies developing policies on the release of outputs to give greater explicit attention to the risks and possible harms that can flow from the uncritical release of *micro* data.

13.6.3 Training

Although not specifically mentioned in the listing of safeguards, training is certainly an important element of any prevention strategy. However, such training will need to go beyond simply the technical aspects of disclosure prevention or the related legal and regulatory requirements for disclosure, as important as such training may be, to cover a broader range of issues, prevention strategies, and safeguards, including ethics.

13.6.4 Further Research

Related to training is the need for further research. This includes further research on known incidents of targeting or attempted targeting to clarify key uncertainties, research into possible additional incidents, several of which have been mentioned earlier in the chapter, meta research into the phenomenon of targeting, and finally research into specific methods and approaches for improving safeguards.

13.6.5 The Establishment of an Incident Register

Our final recommendation takes up an idea mentioned earlier in the chapter when we referred to one shortcoming of Table 13.1, noting that it excluded

what we suspect are numerous unreported incidents where government statistical agencies in the United States and elsewhere have quietly and successfully resisted explicit efforts to use the statistical system to target individuals and vulnerable population subgroups

and suggested that a register of such incidents be developed. Such a register could serve both as a deterrent against efforts aimed at undermining statistical confidentiality and as a record of the success of federal statistical agencies in resisting efforts to compromise their responsibilities to data providers. Such a register might be

organized on an agency basis, with a master register maintained by the Office of Statistical Policy, and with some provision for periodic public accountability, either through an independent intermediary, such as the Committee on National Statistics (CNSTAT) or the Government Accountability Office, or by releasing the information directly to the public. The existence of such a register might not have prevented the gravest of the incidents listed in Table 13.1, but it might have helped to deter some of those that took place in democratic and open societies.

Acknowledgment Note this chapter is a revised and updated version of a paper, “Government Statistics and Individual Safety: Revisiting the Historical Record of Disclosure, Harm, and Risk,” originally prepared for presentation at a workshop, Access to Research Data: Assessing Risks and Opportunities, organized by the Panel on Confidential Data Access for Research Purposes, Committee on National Statistics (CNSTAT), Washington, DC, October 16–17, 2003. We would wish to thank the Panel on Confidential Data Access for Research Purposes and its staff for commissioning the earlier paper and for the comments it provided at the October 2003 Workshop. We would also like to thank S Fienberg, J Gardenier, CL Kincannon, J Nobel, and J Weil for their comments on that paper. We also acknowledge with appreciation the comments provided by two anonymous reviewers. The findings and views expressed in the chapter are solely the responsibility of us.

References

- Aly G, Roth KH (1984) *Die restlose Erfassung: Volkszählen, Identifizieren, Aussondern im Nationalsozialismus* (Complete registration: population census, identity, and separation under national Socialism). Rotbuch Verlag, Berlin
- Aly G, Roth KH (2004) *The Nazi census: identification and control in the third Reich*. Temple University Press, Philadelphia
- Anderson M, Seltzer W (2005) Federal statistical confidentiality and business data: twentieth century challenges and continuing issues. Paper presented at the 2005 Federal Committee on Statistical Methodology (FCSM) Research Conference, Arlington
- Anderson M, Seltzer W (2007) Challenges to the confidentiality of U.S. Federal Statistics, 1910–1965. *Journal of Official Statistics* 23:1–34
- Barabba V (1975) The right of privacy and the need to know. In U.S. Census Bureau, *The census bureau: a numerator and denominator for measuring change*, Technical Paper 37. US Government Printing Office, Washington
- Barabba V (1980) Letter to Raymond Okamura (December 16, 1980). Quoted in part in Okamura RY (1981) The myth of census confidentiality. *Amerasia Journal* 8:111–120
- Barron Jr WG (2002) Email to W Seltzer (January 8, 2002)
- Beeger W, de Vries WFM (1987) Public confidence in official statistics. *Netherlands Official Statistics* 2(3)
- Black E (2001) *IBM and the holocaust: the strategic alliance between Nazi Germany and America’s most powerful corporation*. Crown, New York
- Blum A (2000) Stalinism and population statistics. Paper presented at the annual meeting of the Population Association of America, Los Angeles
- Bohme FG (1975) Letter to Roger Daniels (December 3, 1975). Records relating to the War History Project (item 156); US Census Bureau, Record Group 29; National Archives Building, Washington
- Bohme FG, Pemberton DN (1991) Privacy and confidentiality in the US censuses—a history. Paper presented at the annual meeting of the American Statistical Association, Atlanta
- Bryant BE, Dunn W (1995) *Moving power and money: the politics of census taking*. New Strategist Publications, Ithaca

- Caplan AL (1992) The doctor's trial and analogies to the Holocaust in contemporary bioethical Debates. In: Annas GJ, Grodin M (eds) *The Nazi doctors and the Nuremberg code: human rights in human experimentation*. Oxford University Press, New York, pp 258–275
- Capt JC (1942) Telegram to Dedrick, Hotel Whitcomb, San Francisco (March 21, 1942). NARA, RG40, Commerce, Entry 1, Box 144, File no. 67104
- Census Advisory Committee (CAC) (1942) Full transcript of meeting of January 9–10. Census Advisory Committee Meeting, January 9 and 10, 1942; General Records of the Census Advisory Committee (item 148); US Census Bureau, Record Group 29; National Archives Building, Washington
- Choldin H (1988) Government statistics: the Conflict between research and privacy. *Demography* 25:145–154
- Clark TC (1972) Interview for the Earl Warren Oral History Project. The Bancroft Library, University of California, Berkeley
- Clausen JA, Conrad HS, Freeman WWK, Hobbs N, Hopkins JW, Watson AN, Likert R (1954) Report of the Ad Hoc Committee on Statistical Standards. *The American Statistician* 8:19–23
- Corcoran TF (1963) On the confidential status of census reports. *The American Statistician* 17: 33–40
- Dedrick CL (1942) Japanese population of Bainbridge Island. Memorandum to Col. Bendetsen (March 21, 1942). Bancroft Library, Banc MSS 67/14c FILM, Reel 12, frames 213 and 215; also in NARA, RG338, WCCA, Unclassified correspondence, Box 15, File 093 “Evacuation of Bainbridge Island”
- Dedrick CL (1973) Hand written note, dated November 29, 1973. General records maintained by Calvert Dedrick, 1935–42 (item 210); US Census Bureau, Record Group 29; National Archives Building, Washington
- Dedrick CL (1981) Testimony of Calvert Dedrick before the Commission on Wartime Relocation and Internment of Civilians, November 3, 1981, NARA, CWRIC Hearings Transcripts, pp 180–181
- Dedrick CL (No date) Hand written note. General records maintained by Calvert Dedrick, 1935–42 (item 210); US Census Bureau, Record Group 29; National Archives Building, Washington
- DeRosier Jr AH (1970) *The removal of the Choctaw Indians*. The University of Tennessee Press, Knoxville
- Duncan GT, Jabine TB, de Wolf VA (eds) (1993) *Private lives and public policies: confidentiality and access of government statistics*. Panel on Confidentiality and Data Access, Committee on National Statistics, National Research Council and the Social Science Research Council, National Academy Press, Washington
- Duncan JW, Shelton WC (1978) *Revolution in United States government statistics, 1926–1976*. Office of Federal Statistical Policy and Standards. US Department of Commerce, Washington
- Dunn Jr ES (1967) The idea of a national data center and the issue of personal privacy. *The American Statistician* 21:21–27
- Ellenberg JH (1983) Ethical guidelines for statistical practice: a historical perspective. *The American Statistician* 37:1–4
- El Badry S, Swanson D (2007) Providing census tabulations to government security agencies in the United States: the case of Arab Americans. *Government Information Quarterly* 24: 470–487.
- Elliot CP (1948) An Indian reservation under General George Crook. *Military Affairs* 12:91–102
- Fay RG (2001) Evidence of additional erroneous enumerations from the person duplication study. Executive Steering Committee on Accuracy and Coverage Policy II (ESCAP II), Report 9. Preliminary version, October 26, 2001
- des Forges A (1999) *Leave none to tell the story*. Human Rights Watch, New York
- Freeman WWK (1952) The statistician and his conscience: discussion (in *Standards of Statistical Conduct in Business and Government*). *The American Statistician* 6:18–20
- Gal S (1993) Diversity and contestation in linguistic ideologies: German speakers in Hungary. *Language in Society* 22:337–359.
- Gibbons JD (1973) A question of ethics. *The American Statistician* 27:72–76
- Goldberg JP, Moye WT (1985) *The first hundred years of the Bureau of Labor Statistics*. US Government Printing Office, Washington

- Grodzins M (1949) *Americans betrayed: politics and the Japanese evacuation*. University of Chicago Press, Chicago
- Habermann H (2006) Ethics, confidentiality, and data dissemination. *Journal of Official Statistics* 22:599–614. Paper initially presented as a president's invited address, International statistical Institute, 55th session, Sydney, April 5–12, 2005
- Hochschild A (1998) *King Leopold's ghost: a story of greed, terrorism, and heroism in colonial Africa*. Houghton Mifflin, Boston
- Holley RH (1938) Confidential nature of individual data returned on population schedules. Memorandum to the Director, dated September 8, 1938. NARA, RG29, Entry 274, Box 56 (Holley), Folder: Austin, W.L.
- Horrell M (1959) Race classification in South Africa – its effects on human beings. South African Institute of Race Relations, Johannesburg
- International Statistical Institute (1986) Declaration of professional ethics for statisticians. *International Statistical Review* 227–247. Also available at <http://isi.cbs.nl/ethics.htm> (accessed September 28, 2006)
- Ireland CT, Hayek L-A, Dalenius T, Hollander S, Mann CR, Marks ES, Pierce JS, Scheuren FJ, Seltzer W (1983) Ethical guidelines for statistical practice: report of the Ad Hoc Committee on Professional Ethics. *The American Statistician* 37:5–6
- de Klerk, W (1998) *The last trek—a new beginning: the autobiography*. Macmillan, London.
- Kraly EP, McQuilton J (2005) The 'protection' of Aborigines in colonial and early federation Australia: the role of population data systems. *Population, Place and Space* 11: 225–250
- Jones JH (1981) *Bad blood: the Tuskegee syphilis experiment*. The Free Press, New York
- Jones CD (2005) *The Role of Population Statistics and the Japanese War Relocation Program*, Unpublished paper, US Census Bureau. Copy provided by author
- Lie E (2002) Numbering the nationalities: ethnic minorities in Norwegian population censuses 1845–1930. *Ethnic and Racial Studies* 25:802–822
- Mamdani M (2001) *When victims become killers: colonialism, nativism, and the genocide in Rwanda*. Princeton University Press, Princeton
- McNeil M (2002) Term paper prepared for the course, "Uses/Abuses of Government Population, Social, and Economic Data." Fordham University, Graduate School for the Arts and Sciences, spring semester
- Methorst HW (1936) The new system of population accounting in the Netherlands. *Journal of the American Statistical Association* 31:719–722
- Methorst HW (1938) The new system of population accounting in the Netherlands. *Journal of the American Statistical Association* 33:713–714
- Mitchell WD (1930) Letter from the Attorney General to the Secretary of Commerce, dated September 29, 1930. NARA, RG40, General Records of the Department of Commerce, Office of the General Council, Subject and Index Files, 1903–1947, Box 157 and RG29, Entry 142, Box 3 Austin, Folder: Attorney General, Department of Justice
- Moore B (1997) *Victims and survivors: the Nazi persecution of the Jews in the Netherlands, 1940–1945*. Arnold, London
- Mule T (2001) ESCAP II: Person Duplication in Census 2000. Executive Steering Committee on Accuracy and Coverage Policy II (ESCAP II), Report No. 20. U.S. Census Bureau, Washington DC
- National Research Council (2001) *Principles and Practices for federal statistical agency*. Second edition. Martn ME, Straf ML, Citro CF(eds), Committee on National Statistics, National Academy Press, Washington
- Office of Management and Budget (1997) Order providing for the confidentiality of statistical information. *Federal Register* 62:35044–35050 (www.whitehouse.gov/OMB/infomag/statpol.html, accessed May 30, 2002, www.uwm.edu/~margo/govstat/integrity.htm, accessed September 28, 2006)
- Presser J (1969) *The destruction of the Dutch Jews*. E.P. Dutton and Co., New York. Originally published in Dutch in 1965 as *Ondergang*.
- Prewitt K (2000) Statement issued on March 24, 2000. Quoted in full in U.S. Census Bureau 2004, p 16

- Qin D (2004) Term paper prepared for the course "Demography, Human Rights, and Ethics." Fordham University, Graduate School for the Arts and Sciences, autumn semester.
- Rémond R et al. (1996) *Le "Fichier Juif"* – Rapport de la Commission Présidée par René Rémond au Premier Ministre. Plon éditeur, Paris
- Rice S (1941) Letter to W.F. Ogburn, dated February 6, 1941. University of Chicago Library, Special Collections, W.F. Ogburn papers. Box 27, Census Advisory Committee; folder 5, Bureau of the Census, correspondence
- Rogers S (1917a) Letter to the Secretary of Commerce, dated June 22, 1917. NARA, RG40, General Records of the Department of Commerce, Office of the General Council, Subject and Index Files, 1903–1947, Box 157
- Rogers S (1917b) Letter to the Solicitor, Department of Commerce, dated June 28, 1917. NARA, RG40, General Records of the Department of Commerce, Office of the General Council, Subject and Index Files, 1903–1947, Box 157, file 3400–124
- Rogers J, Nelson MG (2003) 'Lapps, Finns, Gypsies, Jews, and Idiots': modernity and the use of statistical categories in Sweden. *Annales de Démographie Historique* 1:61–79
- Rowe Jr J (1941) Memorandum to FDR, dated February 20, 1941 FDR Library, James Rowe Jr. Collection, Box #5, folder "Census Bureau"
- Seltzer W (1994) Politics and statistics: independence, dependence, or interaction? Working paper No. 6. United Nations Department of Economic and Social Information and Policy Analysis, New York
- Seltzer W (1998) Population statistics, the Holocaust, and the Nuremberg trials. *Population and Development Review*, 24:511–552
- Seltzer W (1999) Excluding Indians not taxed: federal censuses and native-American in the 19th century. Paper presented at the 1999 Joint Statistical Meetings, Baltimore (www.uwm.edu/margo/govstat/integrity.htm, accessed September 28, 2006)
- Seltzer W (2001) U.S. federal statistics and statistical ethics: the role of the American Statistical Association's ethical guidelines for statistical practice. Paper based on presentation to the Washington Statistical Society, February 2001 (www.uwm.edu/margo/govstat/integrity.htm, accessed September 28, 2006)
- Seltzer W (2003) Data collection, ethics issues in. In: Demeny P, McNicoll G (eds) *Encyclopedia of population*. Macmillan Reference USA, New York, pp 195–197.
- Seltzer W (2005) Statistics and counterterrorism: the role of law, policy and ethics. Proceedings of the American Statistical Association section on risk analysis. American Statistical Association, Alexandria, pp 4052–4059
- Seltzer W, Anderson M (2000) After Pearl Harbor: the proper role of population statistics in time of war. Paper presented at the annual meeting of the Population Association of America, Los Angeles (www.uwm.edu/margo/govstat/integrity.htm, accessed September 28, 2006)
- Seltzer W, Anderson M (2001) The dark side of numbers: the role of population data systems in human rights abuses. *Social Research* 68:481–513 (www.uwm.edu/margo/govstat/integrity.htm, accessed September 28, 2006)
- Seltzer W, Anderson M (2002) NCES and the Patriot Act: an early appraisal of facts and issues. Paper presented at the Joint Statistical Meetings, New York, August 10–15 (www.uwm.edu/margo/govstat/integrity.htm, accessed September 28, 2006)
- Seltzer W, Anderson M (2007) Census confidentiality under the Second War posers Act(1942–1947). Paper presented at the annual meeting of the population Association of America, New York, March 30 2007 (www.uwm.edu/margo/govstat/integrity.htm, accessed August 17, 2007)
- Søbye E (1998) Et moerkt kapittel i statistikkens historie? (A dark chapter in the history of statistics?) *Samfunnspeilet* 8(4)
- Stauber R (1946) Final report of the relocation planning division. War relocation authority, type-written report, dated April 26, 1946. Bancroft Library, Banc MSS 67/14c FILM, Reel 26, frames 50ff
- Steuart W (1930) Memorandum to the Secretary of Commerce, dated July 8, 1930. NARA, RG40, General Records of the Department of Commerce, Office of the General Council, Subject and Index Files, 1903–1947, Box 157

- Thomas DS (1937) Review of De Bevolkingsboekhouding by J.L. Lentz. *American Sociological Review* 2:117–118
- Union of South Africa (1950) Population Registration Act. Sections 1–24 in Act No. 30 of 1950
- Union of South Africa Bureau of Census and Statistics (1959) Annual report of the Statistics Council, incorporating the annual report of the Director of Census and Statistics. Government Printer, Pretoria
- Union of South Africa Bureau of Census and Statistics (1960) Union statistics for fifty years, 1910–1960: jubilee issue. Government Printer, Pretoria
- Union of South Africa Parliament House of Assembly (1950a) Debates of the House of Assembly (Hansard). Third Session – Tenth Parliament. January 20 to June 24. Unie-Volkspers Bpk., Cape Town
- Union of South Africa Parliament House of Assembly (1950b) Report. Select Committee on the Population Registration Bill. Chairman T. E. Donges. Cape Times Ltd., Cape Town
- Union of South Africa Parliament Senate (1950) Debates (Official Report). Third Session – Tenth Parliament (Fifth Senate). Unie-Volkspers Bpk., Cape Town
- U.S. Census Bureau (1942a) Exhibit A – Work done on reimbursable basis by Bureau of the Census for war or other agencies charged with execution of some phase of the war program, December 1941 through November 1942. December 4, 1942, mimeographed. War Work Reports folder; National Defense; Papers of Philip Hauser (item 146); US Census Bureau, Record Group 29; National Archives Building, Washington DC
- U.S. Census Bureau (1942b) U.S. Bureau of the Census, Sixteenth Census of the United States, 1940, Population and Housing, Statistics for Census Tracts, Los Angeles-Long Beach, California. GPO, Washington DC
- U.S. Census Bureau (1982) Statement on Census Bureau actions at the outset of World War II as Reported in Infamy: Pearl Harbor and its aftermath by John Toland (Attachment to letter from Bruce Chapman to Raymond Okamura, dated October 21, 1982). CWRIC microfilm
- U.S. Census Bureau (2001) The public's privacy – no idle pledge! U.S. Census Bureau, Washington DC (www.census.gov, accessed October 25, 2001)
- U.S. Census Bureau (2004) Census confidentiality and privacy, 1790–2002 (www.census.gov/prod/2003pubs/conmono2.pdf, accessed September 28, 2006)
- U.S. Congress, Senate, Committee on Military Affairs (1942) S. 2352. Unprinted hearing report prepared by Ward and Paul, official reporters, 77th Congress, 2nd Session, March 13, 1942, pp 8–9. Bancroft Library, Banc MSS 67/14c FILM, Reel 4, frame 69
- U.S. Department of Commerce (1917) IN RE authority of the Director of the Census to furnish registration officials with names and ages of certain persons from census records. Opinion of the Acting Solicitor, Office of the Solicitor, Department of Commerce, dated June 25, 1917. NARA, RG40, General Records of the Department of Commerce, Office of the General Council, Subject and Index Files, 1903–1947, Box 157, file 3400–124
- van de Kaa D (1998) A brief note on population issues in the Netherlands during the 20th century. In: Mackensen R (ed) *Bevölkerungsfragen auf Abwegen der Wissenschaften: Dokumentation des 1. Colloquiums zur Geschichte der Bevölkerungswissenschaft in Deutschland im 20. Jahrhundert*. Leske+Budrich, Opladen, 107ff
- van der Laan P (2000) The 2001 census in the Netherlands integration of registers and surveys. Paper prepared for the conference on 'The Census of Population: 2000 and Beyond' organized by the Cathie Marsh Centre for Census and Survey Research, Faculty of Economics and Social Studies, University of Manchester, Manchester, 22–23 June (www.ccsr.ac.uk/conference/VanderLaanpap.doc, accessed September 28, 2006)
- Wartime Civil Control Administration (1942a) Japanese population and number of Japanese family heads in Arizona, by Minor Civil Division: 1940. Bulletin No. 2. Western Defense Command and Fourth Army, Wartime Civil Control Administration, Statistical Section, March 16, 1942. Bancroft Library, Banc MSS 67/14c FILM, Reel 12, frames 543ff
- Wartime Civil Control Administration (1942b) Japanese population and number of Japanese family heads in California, by Minor Civil Division: 1940. Bulletin No. 3. Western Defense Command and Fourth Army, Wartime Civil Control Administration, Statistical Section, March 16, 1942. Bancroft Library, Banc MSS 67/14c FILM, Reel 12, frames 544ff

- Wartime Civil Control Administration (1942c) Japanese population and number of Japanese family heads in Washington, by Minor Civil Division: 1940. Bulletin No. 4. Western Defense Command and Fourth Army, Wartime Civil Control Administration, Statistical Section, March 16, 1942. Bancroft Library, Banc MSS 67/14c FILM, Reel 12, frames 552ff
- Wartime Civil Control Administration (1942d) Japanese population and number of Japanese family heads in Oregon, by Minor Civil Division: 1940. Bulletin No. 5. Western Defense Command and Fourth Army, Wartime Civil Control Administration, Statistical Section, March 16, 1942. Bancroft Library, Banc MSS 67/14c FILM, Reel 12, frames 557ff
- Wartime Civil Control Administration (1943) Section III miscellaneous statistics. In Bulletin 12. Western Defense Command and Fourth Army, Wartime Civil Control Administration, Statistical Division. March 15, 1943. Bancroft Library, Banc MSS 67/14c FILM, Reel 13, frames 151ff
- Wietog J (2001) *Volkszählungen unter dem Nationalsozialismus. Eine Dokumentation zur Bevölkerungsstatistik im Dritten Reich*. Duncker and Humblot, Berlin

Author Index

- Adlakha, A.L., 23, 25–27
al-Khalil, S., 46
Alonso, J.F., 74
Alston, P., 12
Aly, G., 311
Amowitz, L., 28
Anderson, M., 29, 205, 213, 273, 275, 280,
 281, 283, 285, 286, 290, 293–296, 299,
 300, 301, 302–303, 305, 306, 308–314,
 316, 320, 321
Anderson, R.L., 188
Applebaum, A., 5
Arkin, W., 58
Armstrong, J.D., 14, 15
Ashagrie, K., 23
Asher, J., 3–29, 49, 82, 241
Axelrod, R., 238–239
Axtell, R.L., 232
- Ball, M.M., 9
Ball, P., 11, 20–21, 28–29, 49, 111–113,
 117–121, 133, 149, 157, 200, 202, 203,
 212, 213, 221
Banks, D., 29, 82, 157, 189, 227
Barabba, V., 280, 290
Barnett, M.N., 11
Barron, W.G., 308
Bass, G.J., 7, 11
Bassiouni, M.C., 39
Baynes, K., 231
Bebr, G., 5
Begeer, W., 290, 311
Bennett, J., 111
Bennett, S., 26
Bergsten, J.W., 23–25
Besson, J.-L., 199
Bilder, R.B., 9
Birmingham, W.B., 23–25, 149
Bishop, Y.M.M., 124
Black, E., 292, 310
- Blum, A., 293, 312
Bohme, F.G., 279–282, 297–299, 303
Bourguignon, F., 260
Brass, W., 22, 27
Bräutigam, D.A., 253
Brogan, D., 137
Brogan, P., 48
Broome, R., 43
Brunborg, H., 118, 203, 215, 219–221
Bryant, B.E., 290–291
Bulmer, M., 24
Burgerman, S.D., 11
Burnham, G., 27
- Campbell, D.T., 252, 261, 262
Caplan, A.L., 289
Capt, J.C., 280, 284, 285, 296, 299–301,
 305, 320
Carleton, D., 10
Carlson, B.A., 23–27
Casley, D.J., 22–26, 149
Cassels, A., 244
Castellan, G., 43
Chander, R., 23
Chapman, A., 29
Choldin, H., 311
Clark, T.C., 58, 303
Claude, R.P., 82, 157, 189
Clausen, J.A., 277
Conibere, R., 11, 13
Corcoran, T.F., 286
Corillon, C., 18
Cornell, S.E., 12
Cribb, R., 120
Cuny, F., 54
- Dadrian, V.N., 44
Danner, M., 10
Daponte, B.O., 28, 51, 53, 55, 58–59, 61, 202
Das, T., 6

- David, I.P., 23
 Dean, R.N., 192
 Dedrick, C.L., 290, 296, 299, 301, 302–305, 307, 319–320
 DeGroot, M.H., 212
 Deichmann, U., 244
 Deming, W.E., 22, 122
 DeRosier, A.H., 296
 Des Forges, A.L., 38, 118, 218, 293, 314
 de Vries, W.F.M., 290
 Donnelly, J., 8–9, 12, 17
 Duflo, E., 258, 263
 Dugard, J., 13
 Duncan, G.T., 29, 318
 Duncan, J.W., 279
 Dunn, W., 290
 Dunn, E.S., 280
 Duplaix, N., 45
 Durant, W., 14, 42

 Eagleton, W., 44
 Eberstadt, N., 74
 El Badry, S., 293, 296, 309
 Elkins, C., 44
 Ellenberg, J.H., 277
 Elliot, C.P., 297
 Elliott, K., 60
 Eltinge, J.L., 24, 149
 Epstein, J.M., 232

 Fay, R.G., 308
 Feldman, D.L., 106
 Felkner, J., 258
 Ferris, R., 245
 Fields, P., 241
 Fienberg, S.E., 29
 Fischer, D.D., 8, 9
 Fitch, D.J., 25, 149, 241, 242, 251, 261–262
 Foote, S.B., 215, 216
 Forster, D., 26
 Forsythe, D.P., 7–8, 10, 12, 14–16
 Freeman, W.W.K., 277
 Frelick, B., 201
 Frey, F.W., 23–24

 Gal, S., 292, 312
 Garfield, R., 61
 Gastwirth, J.L., 199
 Gibbons, J.D., 277
 Gibsson, J.L., 11
 Glendon, M.A., 6–7
 Gloss-Soler, S., 70
 Goldberg, J.P., 279
 Goldman, N., 23, 26
 Gordon, A.M., 74

 Gourevitch, P., 11
 Gouws, A., 11
 Grechanovsky, E., 181, 187
 Green, R., 6
 Grodzins, M., 318
 Gunn, G., 120
 Gutman, R., 8, 48

 Habermann, H., 296, 299, 309
 Hagan, J., 13
 Haglund, W.D., 207
 Haines, A., 244
 Halberstadt, S., 243
 Hayner, P.B., 11
 Heckman, J., 252
 Henkin, L., 9
 Herrera, J., 163
 Heuveline, P., 118
 Hoaglin, D.C., 210
 Hoff, P., 232, 234
 Holley, R.H., 281–282, 299
 Hopkins, J.W., 214
 Horrell, M., 316
 Hull, T., 134
 Humana Ch., 65
 Huntington, D., 23
 Hyman, H.H., 26

 Ireland, C.T., 277

 Jabine, T.B., 24, 26, 29, 82, 154, 157, 181–182, 189–191
 Jackson, M.O., 218, 231
 Jahoda, G., 23–25, 149
 James, D.H., 44
 Jardine, M., 11
 Javeline, D., 142, 145
 Jennings, P.J., 27–28, 127
 Johannsen, R.C., 17
 Jones, J.H., 284, 289

 Kadane, J.B., 62
 Kahneman, D., 236
 Kaye, M.E., 10
 Keck, M.E., 8, 16–17
 Kiaer, A.N., 22
 Kiernan, B., 44, 119, 122, 184
 Kipnis, V., 181
 Kish, L., 22, 26, 147, 149
 Knack, S., 253
 Kobrak, P., 106
 Kraly, E.P., 292, 297, 313
 Kremer, M., 258
 Kunz, J.L., 8
 Kuper, L., 41

- Lago, A., 74
 Lamuro Col, R., 60
 Lauriat, P., 23
 Lee, K., 244
 Lemeshow, S., 136
 Levy, P.S., 136
 Lie, E., 281, 292, 312
 Likert, R., 144
 Little, R.J.A., 23
 Lokshin, M., 252, 256
 Lury, D.A., 24–26, 149
- Madigan, F.C., 23, 26
 Mamdani, M., 297
 Marks, E.S., 20, 202
 Mayall, J., 12
 McCormick, J.M., 10
 McDougal, M.S., 5, 11
 McNeil, M., 293, 314
 McQuilton, J., 292, 313
 Meernick, J., 10
 Meier, P., 211
 Mejia, R., 28
 Mesa Lago, C., 74
 Metagora, 23, 28–29, 157–173
 Methorst, H.W., 310, 312
 Minkel, J.R., 258
 Mironko, C.K., 40
 Mitchell, N.J., 10
 Mitchell, R.E., 23–25
 Mitchell, W.D., 282, 283
 Molina Ro, 248
 Moore, B., 310
 Mosteller, F., 211
 Moye, W.T., 279
 Muhsam, H.V., 190–191
 Mule, T., 308
 Murthy, M.N., 22, 24, 25
- Nardin, T., 9
 Neier, A., 10
 Nelson, M.G., 317
 Neuffer, E., 11
 Neyman, J., 22
 Nolde, O.F., 8
- O’Balance, E., 45
 O’Leary, C., 46
 Omran, A.R., 23, 26
 Orth, R., 39
- Palloni, A., 13
 Pemberton, D.N., 279–282, 297–299, 303
 Pereira da Silva, L.A., 263
 Petre, P., 59
- Piquet Carniero, L., 94
 Poe, S.C., 10
 Polen, S., 242, 258
 Porpora, D.V., 10
 Porter, J.N., 38
 Power, S., 52
 Presser, J., 310
 Prewitt, K., 309
 Pritchett, L., 254
- Qin, D., 292, 313
- Rabin, M., 236
 Rao, V., 253
 Rawls, J.B., 231–232
 Reed, H., 145
 Regan, P.M., 10
 Reinikka, R.S., 253
 Reisman, W.M., 5
 Rémond, R., 292, 310
 Resnikoff, I., 187
 Rice, S., 283
 Rieff, D., 8, 48
 Roberts, L., 27, 132
 Robertson, B., 199
 Robertson, G., 9
 Roberts, S., 200
 Rogers, J., 317
 Rogers, S., 280, 297, 298
 Romeu, J.L., 29, 65, 70, 72
 Rosenbaum, P.R., 260
 Roth, K.H., 311
 Roumani, M.M., 43
 Rowe, J., 284
 Rubin, D.B., 260
 Rummel, R.J., 43, 44, 48
- Sabagh, G., 23
 Sach, J.D., 242, 243
 Sachs, J., 236
 Saeedpour, V.B., 46
 Salama, P., 119
 Sallans, B., 232
 Samuelson, D., 29, 82, 181, 187, 189, 192, 203
 Savage, I.R., 188–189
 Scheuren, F.J., 20–21, 25, 29, 82, 118, 122, 141, 149, 192, 241, 242, 258, 261
 Schwarzkopf Gen, H.N., 59
 Schwelb, E., 9
 Scoble, H.M., 16
 Scott, C., 23
 Seber, G.A.F., 127
 Sekar, C.C., 22, 122

- Seltzer, W., 118, 195, 205, 213, 218, 273, 275, 277, 280, 281, 283, 285, 286, 290, 292, 293–294, 295, 296, 299, 300, 301, 302, 303, 305, 306, 308, 309, 310, 311, 312, 313, 314, 316, 318, 320, 321
- Semb, A.J., 13
- Shelton, W.C., 279
- Shue, H., 9
- Sikkink, K., 8, 16–17
- Silva, R., 11, 21, 28, 29, 117, 119, 133
- Simpson, J., 111
- Sivard, R.L., 45
- Slater, J., 9
- Smith, N., 253
- Snow, C.C., 89
- Snow, R.W., 26
- Søbye, E., 292, 310
- Sohn, L.B., 8
- Som, R.K., 23
- Spiegel, P.B., 119
- Spirer, Hi, F., 20, 112, 113, 118, 157, 189, 192, 195, 203, 214, 219
- Spirer, L., 20, 82, 182, 192, 214, 219
- Stahl, M., 127
- Stannard, D.E., 43
- Stauber, R., 319–320
- Steadman, D.W., 207
- Stettinius, E.R., 7
- Steuart, W., 282–283
- Stohl, M., 10
- Stover, E., 182, 184, 189
- Sudetic, C., 11
- Sudman, S., 127
- Swanson, D., 293, 296, 309
- Swiss, S., 27, 28
- Tabeau, E., 203, 219, 221
- Tashjian, J.H., 44
- Tezcan, S., 23, 26
- Thakur, R., 16, 17
- Thomas, D.S., 310
- Thompson, S.K., 127
- Tufte, E., 203
- Tukey, J.W., 210, 211
- Ubelaker, D.H., 43
- Urdal, H., 219
- Vaessen, M., 23
- van Bruinessen, M., 46
- van de Kaa, D., 312
- van der Laan, P., 311
- Velleman, P.F., 210
- Verma, V., 23, 26
- Verpoorten, M., 219
- Verwimp, P., 40, 203, 219
- Vigneaux, G.A., 199
- Waddingham, J., 120
- Wall, B., 80, 245
- Walzer, M., 52, 54
- Warwick, D.P., 24
- Wassenich, P., 241
- Wassink, J., 43
- Watkin, K., 13
- Weissbrodt, D., 15
- Wellbank, J.H., 231
- Werful, F., 5
- Whiteman, M.M., 7
- Whitmore, T.M., 42
- Wietog, J., 311
- Wilkie, J.W., 78
- Williams, J., 200
- Wiseberg, L.S., 16
- Wolfson, L., 62
- Wood, S., 243
- Woolcock, M., 253
- Wright, C.R., 279, 282
- Yang, D., 5
- Yansaneh, I.S., 24, 149
- Yemtsov, R., 252, 256

Subject Index

- Afghan refugee camps in Pakistan, 141
Africa, 244–245
Agent-based model of human rights,
234–235, 239
Agree–disagree approach, 145
Altruism, 236–237
American Association for the Advancement of
Science (AAAS), 157, 183–192
American Association for the Advancement
of Science’s Science and Human
Rights Program (AAAS-SHRP),
19–20
American Statistical Association (ASA), 82,
181–192
ethical guidelines, 277–278
Amnesty International (AI), 15–16, 191
Anfal, the, 46–47
Arab-Americans, 309
Argentina, 182
“dirty war”, 182
Arusha Tribunal, 40–41
Asia, 245
Australia, aborigines, 313
Axelrod, R., 238–239
- Ball, Patrick, Dr., 20–21, 221
Barankin, Edward, W., 183
Barzani killings, 46–47
Bayesian Demographic approach, 55–56
Begun petition, 191
Bolivia, 246
Boomerang pattern, 17
Bosnia, genocide in, 47–48
Brunborg, Helge, 219–221
- Cambodia, genocide in, 47
Capt, J.C., 284–285, 299–301
Cedula, 90–91
Census Bureau, 274, 280, 283–286, 291,
296–301, 309, 319
- “Chemical Ali”, 46
Chen, Wen-chen, 184, 186
Chile, 11, 71–72
longitudinal comparison, 73, 76
socioeconomic indicators, 72, 77
China, Cultural Revolution, 313
Civil Society Organizations (CSOs), 167
Commission on Reception, Truth and
Reconciliation (CAVR), 120–121
Committee on International Relations in
Statistics, 182
Committee on Scientific Freedom and Human
Rights, 181–191, 230
cases from Soviet Union, 187
charge of, 185
new draft, 189
establishment of Ad Hoc Committee,
182–185
establishment of Committee, 182–183
Confidential Information Protection,
286–287
Confidentiality of information, *see* Statistical
confidentiality
Consumer Price Index (CPI), 69
Convention on the Prevention and Punishment
of the Crime of Genocide, 38
Costa Rica, 71–72
longitudinal comparison, 73, 76
socioeconomic indicators, 72, 77
Cost-benefit analysis of social programs,
235–237
Covenant on Civil and Political Rights,
8–9, 191
Covenant on Economic, Social, and Cultural
Rights, 8–9, 191
Cuba, 66–67
as case study for human rights violations,
71–81
longitudinal comparison, 73, 76
longitudinal study approach, 79–81

- Cuba (*cont.*)
 pre-/post-Test comparisons, 74–75
 socioeconomic indicators, 72, 77
- Darfur crisis, 45
- Data collection in developing world, 24–26
 advantages, 25
 analysis of, 27
 indirect estimation, 27
 issues, 24–25, 68, 241–242
 potential solutions, 25–26
- Data transparency, 253
- Data, types of, 275
 macro, 275, 294, 301–302
 meso, 275, 290–291, 294, 302–303
 micro, 275, 279–282, 290–291, 294,
 305–306
- Dedrick, C.L., 301–302
 memorandum, 303–304, 319–320
- DEFT, 146–148
- Democide, 48
- Demographic analysis of war, 53
- Desaparecido, 89–90, 108, 112–113
- Difference Principle, 231
- Differential increments, 65, 70, 75–78
- Direct conflict deaths, in Timor-Leste, 132–133
- Disclosure issues, 290
 barriers to study of, 317–320
See also Statistical confidentiality
- Donors, 255–259
 Inter-American Development Bank (IDB),
 256–257
 Millennium Challenge Corporation,
 257–259
 USAID, 257
 World Bank, 255–256
- Dunn, Edgar, Jr., 280
- Durant, Henri, 14
- Economic sanctions, 11, 60–61
- El Rescate, 19–20
- El Salvador, 19–20, 188
- Ethical Guidelines for Statistical Practice, 277
- Ethics, 276–278
- Ethics Declaration, 277–278
- Excess mortality, in Timor-Leste, 134–135
- Exploratory Data Analysis (EDA), 210
- Extractive Industries Transparency
 Initiative, 253
- Extradition, 13
- Federal Statistical Confidentiality Order,
 280–281
- Fieldwork, 216
- Fifteenth Census, 282
- Fifteenth Census Act, 282
- Fourteenth Census, 282
- France, 311
- Genocide
 casualties count, 41–42
 methods of calculating, 43–44
 definition of, 38–41
 history of, 42–43
 statistics of, 37–49
- Germany, 311
- Global advocacy network, 17
- Global human rights regime, 17
- Government statistical agencies, 274
 Census Bureau, 274, 280, 283–286, 291,
 296–301, 309, 319
- Graveyard Census Database (GCD), 121–122
 strengths and limitations, 124–127
- Grechanovsky, Eugene, 181
- Greene, Samuel, 184–185
- Grudges, 237–239
- Guatemala, 20–21, 66–67, 89–116, 118, 246
 civil war, 66–67
 International Centre for Human Rights
 Research, 20–21
 medicolegal death investigation in, 90–91
 state terrorism in, 105–109
 Lucas Garcia government, 105–115
 rural campaign, 105
 urban campaign, 105–110
 violent deaths in, 97–116
 during BASE period, 93, 100
 during ELEV period, 93, 99–109
 X.X.s, 89–116
 burial, 90–95
 burial certificates, 93
 and guerilla deaths, 107
 natural deaths, 97
 violent deaths, 97–100
- Hassan al-Majid, Ali, *see* “Chemical Ali”
- Holocaust, 310
 in France, 311
 in Germany, 311
 in the Netherlands, 311–312
- Holocaust-related examples, of misusing
 population data systems to target
 subgroups, 310
 in France, 311
 in Germany, 311
 in the Netherlands, 311–312
- HR, *see* Human Rights (HR)
- Humanitarian advocacy, 5
- Humanitarian law, *see* International
 humanitarian law

- Human rights (HR), 65–84
 abuses, 4–5, 11, 18–19, 28, 202–206, 213
 data modeling and analysis, 66
 challenges of, 68–70
 definition of, 3–4
 history of, 228–231
 law, history of, 5–13
 meaning of, 65
 problem statement of, 65–68
- Human rights, agent-based model of, 234–235
- Human rights abuses, 4–5, 11, 18–19, 28, 204–209, 213
- Human Rights Accountability Coalition (HRAC), 167
- Human Rights Information and Documentation Systems (HURIDOCS), 19
- Human rights institutions (HRIs), 165–166
- Human rights law, history of, 5–13
See also International human rights law
Human Rights Quarterly, 189
- Human rights violations, 65–84
 body-of-knowledge, 67
 collateral damage, 67
 Cuba as case study, 71–81
 Cuban revolution, 66–67
 data problems, 69–70
 differential increments in country
 comparison, 70, 75–78
 justification of, 83–84
 problem statement of, 65–66
 statistical analysis of, 65–84
- Human Rights Violations Database (HRVD), 120–134
 strengths and limitations, 124–127
- Hungary, 312
- Huridocs Standard Formats: A Tool for Documenting Human Rights Violations*, 19
- Hussein, Saddam, 46
- Impact evaluations, 252–255
 qualitative techniques, 252–253
 quantitative techniques, 252–253
 reasons for lack of, 254
- Indirect conflict deaths, in Timor-Leste, 132–133
- Information management systems, 212
- Input-output model, 236–237
- Institutional incentives, 259
- Inter-American Development Bank (IDB), 256–257
- International Criminal Court (ICC), 195–223
 attributes of, 197
 civil law tradition, 217
 establishment of, 196
 role of statistics and demography, 198–217
- International Criminal Tribunal, 198, 200
- International Criminal Tribunal for Former Yugoslavia (ICTY), 198, 217–222
 common law tradition, 217–218
- International Criminal Tribunal for Rwanda (ICTR), 198, 218–219
 common law tradition, 217–218
- International humanitarian law, 4
 Geneva Conventions, 4–5
 and humanitarian advocacy, 5
- International human rights law, 5
 and human rights advocacy, 5
See also Human rights law, history of
- International human rights treaties, 3–4, 188
 civil and political rights, 3–4
 economic, social and cultural rights, 4
 violations of, 4
- International Red Cross, 14–15
 International Committee of the Red Cross, (ICRC), 14–15
 League of Red Cross societies, 14–15
 national Red Cross societies, 14–15
- International Statistical Institute, 182, 190–191
 ethical guidelines, 277–278
- Investment, 235–237
- Iraq, 11
 mass killings, 46–47
 U.S. invasion of, 13
- Iraq War (2003), 60
- Iterated Prisoner's Dilemma, 238–239
- Jackson, Robert, 218
- Jamaica, 246
- Japanese American experience, 299–307, 319–320
- Just and Unjust Wars*, 52
- Kaufman, Daniel, 169, 172
- Khmer regime, 47, 118
- Kipnis, Victor, 181
- Kuwait, 11
- Latin America, 246
- La Verbena, cemetery of, 90
- League of Nations, 6
- Lemkin, Raphael, 8
- Leone, Fred, 182–183, 191
- Liberty Principle, 231
- Lot Quality Assurance Sampling, Valedéz's, 246–249

- Mellibovsky, Graciela, 184
- Metagora, 157–173
 - activities of, 159–160
 - as decentralized laboratory, 158–160
 - future work, 172–173
 - global and national indicators of governance, 171–173
 - governance, democracy and human rights, 171–172
 - initiatives, 168
 - land reform in South Africa, 161
 - major accomplishments of, 161
 - questionnaire of initiatives measuring democracy, human rights, and governance, 174–178
- Metagora, activities
 - in Mexico, 162
 - in Palestine, 160
 - in South Africa, 162
 - in the Philippines, 162, 166
- Metagora Forum, 158, 172
- Mexico, 71–72
 - longitudinal comparison, 73, 76
 - Metagora activities, 161–162
 - socioeconomic indicators, 72, 77
- Millennium Challenge Corporation, 257–259
- Millennium Development Goals, 242–243
- Milošević, Slobodan, 12, 20–21
- Mortality
 - due to postwar chaos, 58
 - from direct effects of war, 57
 - from indirect effects of war, 55–57
- Mortality in Timor-Leste, conflict-related
 - challenges in measuring, 119–120
 - data sources used for estimation, 124–135
 - comparative analysis of, 124–135
 - Graveyard Census Database (GCD), 121–122
 - Human Rights Violations Database (HRVD), 120
 - Multiple Systems Estimation (MSE), 118, 122–124
 - Retrospective Mortality Survey (RMS), 122
- Multiple Systems Estimation (MSE), 118, 122–135
 - capture-tag-recapture, 123–124
- National Statistical Offices (NSOs), 163–164
- Native Americans, 296–297
- Needless Deaths*, 58
- Negreros, Liliana, 107
- Nepal, 245
- Netherlands, 311–312
- NGOs, *see* Nongovernmental Organizations (NGOs)
- Nigeria, 244
- Ninth Census, 279
- Nongovernmental organizations (NGOs), 13–17
 - Amnesty International (AI), 15–16
 - International Red Cross, 14–15
 - International Committee of the Red Cross (ICRC), 14–15
 - League of Red Cross societies, 14–15
 - national Red Cross societies, 14–15
 - role of, 16–17
 - enforcement, 16–17
 - monitoring, 16–17
 - norm generation, 16
- Norgle, Charles, R., 238
- Noriega, Carlos, 182–184, 191
- Norway, 312
- Nuremberg code, 277
- Nuremberg Trial, 290
- Nuremberg Tribunal, 198, 218
- Office of the High Commissioner for Human Rights (OHCHR), 12
- Oportunidades, 251–252
- Overkill, 100
- Panel on Statistical Assessments as Evidence in the Courts, 199, 214
- Partner Organizations, 159–160
- Patriot Act, 308
- Persian Gulf War (1991), 58–59
 - and Rule of Proportionality, 59
- Pinsker, Ilya, 181
- Population data systems, misuse in targeting subgroups and individuals, 291–320
 - Arab-Americans, 309
 - Australia, aborigines, 313
 - China, Cultural Revolution, 313
 - other European examples, 312
 - Hungary, 312
 - Norway, 312
 - USSR, former, 312–313
 - Holocaust-related, 310
 - in France, 311
 - in Germany, 311
 - in the Netherlands, 311–312
 - Japanese American experience, 299–308, 318–319
 - role of Census Bureau staff, 303
 - role of macro data, 301–302
 - role of meso data, 302–305
 - role of micro data, 305–308

- Native Americans, 296–297
- Patriot Act, 308
- Rwanda, 313–314
- South Africa, apartheid, 314–316
- World War I draft registration prosecutions, 297–299
- Poverty, 242–243
- Poverty Action Lab, 258
- President Taft, 279, 282
- Prevention strategies, against misuse of
 - population data systems to target subgroups, 321
 - attention to meso data risks, 322
 - attention to past factors responsible for misuse, 319
 - establishment of incident register, 322–323
 - further research, 322
 - training, 322
- Primary sampling unit (PSU), 143
- Principle of nonintervention, 9, 13
- Proclamation of Teheran, 9
- Program evaluations, *see* Impact evaluations
- Progres, 251–252
- Propensity Score Matching (PSM), 261
- Prosecutor v. Radislav Krstic*, 220
- Prosecutor v. Slobodan Milošević*, 221
- Public graveyard, definition of, 121
 - marked, 121–122
 - unmarked, 122
- Qualitative human rights violation data, 17–19
 - into quantitative data, 19
- Randomization, 258, 261
- Random sample surveys, 21–27
 - history of, 22–23
 - on human rights violations, 27–28
- Rapaport, Anatol, 239
- Rawls, John, 231
- Reagan, Ronald, 10
- Reference Manual on Scientific Evidence: For Use with Federal Practice and Procedure*, 199
- Resource expenditure and result assessment, problems in, 243
- Retrospective Mortality Survey (RMS), 122–137
 - strengths and limitations, 124–127
 - weight adjustment calculations, 137
- Risk disclosure, 289–291, 317–320
- Role of statistics and demography, at the International Criminal Court (ICC), 198–217
 - cost of statistical and demographic work, 215–216
 - education and training of ICC workers, 212–213
 - ethical standards, 212–213
 - methods of statistical analysis, 210–212
 - descriptive statistics, 210–211
 - inferential statistics, 211–221
 - processing and compilation of data, 212
 - role of sampling, 208–210
 - roles of lawyers, investigators, scientists, and the Court, 216–217
 - sources of data, 204–208
 - administrative registers, 205–207
 - case studies, 208
 - censuses, 204–205
 - medical and anthropologic forensic studies, 207
 - sample surveys, 205
 - standards for assessing evidence, 216
 - types of data and methods, 199–200
- Roosevelt, Eleanor, 7
- Roosevelt, Theodore, 6
- Roosevelt Administration, 283
- Rule of Proportionality, 51–62
 - and casualty estimates, 51–54
 - and economic sanctions, 60–61
 - and Persian Gulf War (1991), 59
- Rwanda, 11
 - genocide in, 37–41, 118, 218, 313–314
- Sachs, Jeffery, D., 242–243
- Samoa, 245
- Santa Cruz massacre, 127
- Savage, Richard, 188–189
- Scientific Freedom and Human Rights (SFHR) Committee, 82
- Second War Powers Act, 285, 306
- Seltzer, William, 118
- Senegal, 244
- Silva, Romesh, 21
- Smart bombs, 60
- South Africa, 9
 - apartheid, 314–316
 - human rights violations, 9
 - Truth and Reconciliation Commission (TRC), 10
- Soviet “refuseniks”, 181
- Statistical and demographic techniques, 118
 - indirect demographic methods, 118
 - Multiple Systems Estimation (MSE), 118, 122–135
 - sample survey techniques, 118
- Statistical confidentiality, 276–289
 - ethics, 276–278
 - risk disclosure, 289–291, 317–320

- Statistical confidentiality (*cont.*)
 statistical policy, 279–281
 statutory protections, 281–289
- Statistical methods, and Millennium
 Development Goals, 241–242
- impact evaluations, 252–255
 qualitative techniques, 252–253
 quantitative techniques, 252–253
 reasons for lack of, 254
- issues, 244, 248
 in Africa, 244–245
 in Asia, 245
 in Latin America, 246
- success stories
 Dominican Republic, 250
 Grenada, 251
 Mexico, 251–252
 Nigeria, 251
 web resources, 262–263
- Statistical policy, 279–281
- Statisticians, roles for, 259–261
- “Statistics, Statisticians and Human Rights”,
 190–191
- Statistics of human rights violations, 21, 28
- Statute of Rome, 196, 212–213
- Statutory protections, 281–289
- Stewart, Ethelbert, 279
- Sudan, 45–46
- Survey of Afghan refugees in Pakistan,
 142–153
 DEFT, 146–147
 purpose and analyses of survey, 142–144,
 149–154
 questionnaire design, 144–145
 sample design, 143
 survey design, 144–145
 survey questions and responses, 151–154
 top refugee concerns, 144
- Tanzania, 245
- Theories of Justice, agent-based, 231–235
 hierarchical probability models,
 233–234
 simulations, 234–235
- Thirteenth Census, 279–280
- Timor-Leste, 117–137
 challenges in measuring, 119
 conflict-related mortality in, 117–137
 data sources used for estimation,
 120–135
 comparative analysis of, 124–135
 Graveyard Census Database (GCD),
 121–122
 Human Rights Violations Database
 (HRVD), 120–121
- Multiple Systems Estimation (MSE),
 118, 122–124
- Retrospective Mortality Survey (RMS),
 121
- Tit-for-Tat, 37
- Treaty of Versailles, 239
- Truth and Reconciliation Commission (TRC),
 10–11
- Twelfth Census, 281–282
- U.S. record, of misusing population data
 systems to target subgroups,
 292–306
 Arab-Americans, 309
 Japanese American experience, 299–308
 role of Census Bureau staff, 299–301
 role of macro data, 301–302
 role of meso data, 302–305
 role of micro data, 305–306
 Native Americans, 296–297
 Patriot Act, 308
 World War I draft registration prosecutions,
 297–299
- Uganda, 10, 245
- UN Commission on Human Rights, 7
- United Nations, 6–7
- United Nations Statistical Commission, 281
- United States, international policy, 10
- United States v. Esquivel*, 274
- Universal Declaration of Human Rights, 7–8,
 191, 228–230
- USAID, 257
- US Centre for Disease Control (CDC), 119
- USSR, former, 312–313
- Utilitarianism, 231
- Utility function model, 237
- Veil of Ignorance, 231
- Vulnerable individuals, 291–294
- Vulnerable population subgroups, 293–297
- Walker, Francis, A., 279
- Walzer, Michael, 52
- War, 51–62
 demographic analysis of, 53–57
 impact on fertility, 53
 impact on mortality, 53–54
 and increased migration, 53
- War Crimes: A Report on United States War
 Crimes Against Iraq*, 58
- Weagers, Margaret, 20
- Web resources, 262–263
- Wellington, Dorothy, 183
- World Bank, 255–256
- Wright, Carroll, 279

- X.X.s, 89–116
 - burial, 111–112
 - burial certificates
 - limitations of data, 93
 - primary variables, 92
 - secondary variables, 92–93
 - and guerilla deaths, 107
 - natural deaths, 97
 - violent deaths, 97–100
 - by presidential regimes, 105–106
- Yugoslavia, 11–12, 119