National Symposium on Family Issues

Susan M. McHale
Paul Amato
Alan Booth
Editors

Emerging Methods in Family Research

Springer

National Symposium on Family Issues

Series Editors
Alan Booth and Susan M. McHale

For further volumes:
http://www.springer.com/series/8381

Susan M. McHale • Paul Amato • Alan Booth
Editors

Emerging Methods in Family Research

Springer

Editors
Susan M. McHale
Department of Human Development
& Family Studies
The Pennsylvania State University
University Park
Pennsylvania
USA

Paul Amato
Department of Sociology
The Pennsylvania State University
University Park
Pennsylvania
USA

Alan Booth
Department of Sociology
The Pennsylvania State University
University Park
Pennsylvania
USA

ISSN 2192-9157 ISSN 2192-9165 (electronic)
ISBN 978-3-319-01561-3 ISBN 978-3-319-01562-0 (eBook)
DOI 10.1007/978-3-319-01562-0
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013949340

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Across a range of disciplines, including psychology, sociology, education, political science and medicine, research and theory target families as central to the well-being of their members and to the well-being of the communities and larger societies in which families are embedded. The significance of families for the health of individuals and communities demands scholars' best efforts to illuminate how family roles, relationships and dynamics operate and how they influence family members.

This volume is predicated on the idea that advances in research on families will rely on innovations in design, measurement, data collection and data analysis that allow researchers to capture the multi-level complexities of family systems. Methods for studying families are often drawn from research focused on individuals. A theme throughout this volume is whether and to what extent the same kinds of methods can be applied across levels of analysis—from the individual, to the dyad, to larger family groups. In chapters throughout this volume, authors consider whether and how methods from research focused on individuals can be applied, can be modified, and are challenged when family relationships and family influences are the focus of study.

The contributions to *Emerging Methods in Family Research* are based on papers presented at the 20th Annual Penn State Symposium on Family Issues held in October, 2012. This edited volume is the culmination of two days of stimulating presentations and discussions organized around four topics: (1) strategies for quantitative analysis of variation and change in families, (2) approaches to analyzing families as systems, (3) measuring "the family" and family dynamics, and (4) new directions in the implementation and evaluation of family-focused social policies and preventive interventions.

Overview of this Volume

This volume is organized by these four topical areas. In contrast to other volumes in the decades-long Family Symposium series, our focus on methods meant that many of these chapters were written by researchers who do not self- identify as family scholars, but rather, are known for their methodological expertise. These methodologists accepted the editors' invitation to apply their work and ideas to the study of families. The four sections of this volume each include two or three chapters that address the topical area in distinct ways, often from different disciplinary perspectives. The last chapter in each section is an integrative discussion by a family scholar who was charged with distilling the range of ideas, information, and techniques described in the session's papers toward providing insights on how novel methods could be used to advance the work of family scholars. The volume concludes with an integrative chapter by two young scholars.

Chapters in Part I focus on best methods for capturing variation and change in family processes and influences on individual family members. Family structure and processes are dynamic, responsive both to changes in individual family member's development, as well as to pressures emanating from outside the family, which are also continually in flux. Chapters by Jay Teachman, Professor of Sociology at Western Washington University, by Nilam Ram, Associate Professor of Human Development at Penn State, and colleagues, and by Si-Miin Chow, Associate Professor of Human Development at Penn State, and colleagues, focus on different timescales for studying variation and change, timescales that reflect different kinds of research questions and require different kinds of analytic methods. In the final chapter in this section, Andrew Fuligni, Professor of Psychiatry and Behavioral Sciences at UCLA, outlines some of the contributions to our understanding of family processes and family influences that can come from sophisticated analyses of variation and change, and he considers how the benefits of collecting "repeated data" balance against the costs. For researchers interested in why one might use the models introduced in the three opening chapters, Fuligni's application of each model to the case of family sleep patterns conveys the distinctive insights that can emerge from each approach.

Family scholars have long embraced the metaphor of families as systems, yet empirical research targeting systems dynamics remains very rare. In Part II, chapters by Robin Gauthier and James Moody, both sociologists at Duke University, and by Mark Cummings and co-authors Kathleen Bergman and Kelly Kuznicki, psychologists at Notre Dame University, focus on methods for characterizing family systems and capturing their dynamics. In his integrative discussion, Robert Emery, Professor of Community Psychology at the University of Virginia, reinforces and elaborates on the important conceptual and theoretical work that must be accomplished if family researchers are to make full use of a systems approach, and he offers new ideas toward this end.

At a general level, measurement is "concerned with what can be observed, the conditions under which observations are made, and how observations are recorded for future analysis and consideration" (Amato, Chap. 11, p. 179). In Part III, chapters by Carolyn Tucker Halpern from the Department of Maternal & Child Health along with Kathleen Mullan Harris from the Department of Sociology and epidemiologist Eric Whitsel, all at University of North Carolina, Chapel Hill, by Joshua Smyth, Professor of Biobehavioral Health and co-author Kristin Herron from Penn State University, and by Thomas Weisner, Professor of Psychiatry and Anthropology at UCLA, describe distinct approaches to measuring family dynamics and their correlates. As with other dimensions of methods considered in this volume, most family research relies on measurement approaches that were developed to study individuals, and these chapters include consideration of approaches for and challenges to moving from the individual to the dyad and group levels of analysis in measuring family processes and influences. In the concluding chapter, Paul Amato, Professor of Sociology at Penn State, considers some of the strengths of these approaches to measurement and provides examples of how each might be applied to address novel questions about family processes and influences. Amato also reminds us of the challenges of defining "the family" in determining strategies for its measurement.

Recent national efforts have been directed at promoting the translation of science to application and practice as well as improving the quality of programs and policy through a focus on evaluation. Although a stronger emphasis on applying research in evidence-based programs and policies is welcome, the development, implementation and evaluation of programs and policies for families face unique challenges. The chapters in Part IV by Carol Metzler, from the Oregon Research Institute and colleagues, by Quinn Moore and Robert Wood from Mathematica Policy Research, and by Linda Collins, Professor of Human Development at Penn State, highlight new approaches to optimal design, implementation and evaluation of the effects of family programs and policies and consider some of the challenges that need to be overcome toward these ends. In the final, integrative chapter of this section, Greg Duncan, from the School of Education at the University of California Irvine, identifies a number of "best practices" in family-focused evaluation and policy research.

As is the tradition in the Family Symposium series, the final chapter of the volume was written by two scholars in the early stages of their careers as family researchers, Melissa Lippold, from Human Development and Family Studies, and Catherine McNamee, from the Population Research Institute at Penn State. Their charge was to bring to bear their distinct disciplinary backgrounds—in human development and demography, respectively—on the ideas and insights conveyed during the four sessions of the conference. Lippold and McNamee identify five themes that cut across chapters in this volume: approaches to defining "family" and capturing its complexities, assessing change and variation in families, the challenges inherent is studying families, the importance of keeping in sight the "big picture," and the significance and special considerations involved in family research that is aimed at improving public health. Lippold and McNamee conclude with their thoughts about opportunities and challenges facing the next generation of family scholars.

Acknowledgments

The editors are grateful to the many organizations at Penn State that sponsored the 2012 Symposium on Family Issues and this resulting volume, including the Population Research Institute; the Children, Youth and Families Consortium; the Prevention Research Center; the Methodology Center; the departments of Sociology, Human Development and Family Studies, Psychology, Biobehavioral Health, Anthropology, and Labor Studies; and the Women's Studies Program. The editors also gratefully acknowledge essential core financial support in the form of a five-year grant from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), R13- HD048150, as well as advice from Regina Bures and Rosalind Berkowitz King of NICHD. The ongoing support of all of these partners has enabled us to attract excellent scholars from a range of backgrounds and disciplines—-the sort of group on whom the quality of the series depends.

A lively, interdisciplinary team of scholars from across the Penn State community meets with us annually to generate symposia topics and plans and is available throughout the year for brainstorming and problem solving. We appreciate their enthusiasm, intellectual support, and creative ideas. In the course of selecting speakers, symposium organizers consult with a wide range of people at other universities, at NICHD, and at other organizations in order to identify highly qualified scholars to participate in the symposium. We also sincerely thank Diane Felmlee, Jennifer Graham, and Wayne Osgood for presiding over symposium sessions.

The efforts of many individuals went into planning the 2012 symposium and producing this volume. We are especially grateful for the assistance of the administrative staff in the Population Research Institute and Social Science Research Institute at Penn State, including Sherry Yocum, Donna Panasiti, Angela Jordan, and Miranda Bair. Finally, neither Symposium nor this volume would have been possible without Carolyn Scott, whose organizational skills, commitment, and attention to the many details that go into developing an engaging conference and an edited book series make it possible for us to focus on the ideas.

Contents

Contributors

Paul Amato, Ph.D. Department of Sociology and Crime, Law, & Justice, The Pennsylvania State University, University Park, PA 16802, USA

Kathleen N. Bergman, M.A. Department of Psychology, Notre Dame University, Notre Dame, IN, USA

Sy-Miin Chow, Ph.D. Department of Human Development and Family Studies, The Pennsylvania State University, University Park, PA, USA

Linda M. Collins, Ph.D. The Methodology Center, Department of Human Development and Family Studies, The Pennsylvania State University, University Park, PA, USA

David Conroy, Ph.D. Department of Kinesiology, The Pennsylvania State University, University Park, PA, USA

E. Mark Cummings, Ph.D. Department of Psychology, Notre Dame University, Notre Dame, IN, USA

Shawna Doerksen, Ph.D. Department of Recreation, Park and Tourism Management, The Pennsylvania State University, University Park, PA, USA

Greg J. Duncan, Ph.D. School of Education, University of California, Irvine, CA, USA

Robert E. Emery, Ph.D. Department of Psychology, Center for Children, Families & the Law, University of Virginia, Charlottesville, VA, USA

Andrew J. Fuligni, Ph.D. Department of Psychiatry & Biobehavioral Science, University of California, Los Angeles, CA, USA

Robin Gauthier Department of Sociology, Duke University, Durham, NC, USA

Carolyn Tucker Halpern, Ph.D. Department of Maternal & Child Health, University of North Carolina Chapel Hill, Chapel Hill, NC, USA

Kathleen Mullan Harris, Ph.D. Department of Sociology, University of North Carolina Chapel Hill, Chapel Hill, NC, USA

Kristin E. Heron, Ph.D. Survey Research Center, The Pennsylvania State University, University Park, PA, USA

Kelly A. Kuznicki Department of Psychology, Notre Dame University, Notre Dame, IN, USA

Melissa A. Lippold, Ph.D. Department of Human Development and Family Studies, The Pennsylvania State University, University Park, PA, USA

Erika S. Lunkenheimer, Ph.D. Department of Human Development & Family Studies, Colorado State University, Fort Collins, CO, USA

Whitney I. Mattson, M.S. Department of Psychology, University of Miami, Miami, FL, USA

Catherine B. McNamee, Ph.D. Population Research Institute, The Pennsylvania State University,University Park, PA, USA

Daniel S. Messinger, Ph.D. Department of Psychology, University of Miami, Miami, FL, USA

Carol W. Metzler, Ph.D. Oregon Research Institute, Eugene, OR, USA

James Moody, Ph.D. Department of Sociology, Duke University, Durham, NC, USA

Quinn Moore, Ph.D. Mathematica Policy Research, Princeton, NJ, USA

Nilam Ram, Ph.D. Department of Human Development and Family Studies, The Pennsylvania State University, University Park, PA, USA

Julie C. Rusby, Ph.D. Oregon Research Institute, Eugene, OR, USA

Matthew R. Sanders, Ph.D. Parenting and Family Support Centre, University of Queensland, Brisbane, QLD, Australia

Mariya Shiyko, Ph.D. Department of Applied Counseling and Educational Psychology, Northeastern University, Boston, MA, USA

Joshua M. Smyth Departments of Biobehavioral Health & Medicine, The Pennsylvania State University, University Park, PA, USA

Jay Teachman, Ph.D. Department of Sociology, Western Washington University, Bellingham, WA, USA

Thomas S. Weisner, Ph.D. Departments of Psychiatry and Anthropology, Los Angeles, CA, USA

Eric A. Whitsel, M.D., M.P.H. Departments of Epidemiology and Medicine in the Gillings School of Global Public Health and School of Medicine, University of North Carolina, Chapel Hill, NC, USA

Robert G. Wood, Ph.D. Mathematica Policy Research, Princeton, NJ, USA

Part I
Family Development and Change

Chapter 1
Latent Growth Curve Models with Random and Fixed Effects

Jay Teachman

The availability of longitudinal data on families is increasingly common. Data sets such as the Panel Study of Income Dynamics, the various National Longitudinal Studies, AddHealth, and Fragile Families allow researchers numerous opportunities to observe and model family-related processes and outcomes as they evolve over time. Accordingly, a number of statistical procedures have been developed to model repeated observations of families and individuals. Two common alternatives are random-effects models (REM) and fixed-effects models (FEM) (Allison 2005; Bollen and Brand 2010) Within the general random-effects framework, latent growth curve models (LGCM) are a useful extension because they allow researchers to explicitly model the *trajectory* of change in an outcome (Lyons and Sayer 2005) In this chapter, using a structural equation modeling (SEM) approach, I demonstrate that LGCMs can also be estimated within a fixed-effects framework. In addition, I show that time-constant covariates, which are generally modeled on the inter-subject level in LGCMs, can be modeled on the intra-subject level. These models are illustrated using data on marital status, education, and body mass index (BMI) for 1761 men taken from four waves (1992, 1996, 2000, 2004) of the 1979 National Longitudinal Survey of Youth (NLSY). Umberson et al. (2009). provide a discussion of the relationship between BMI, marital status, and other covariates. Finally, I show that LGCMs can be used to model paired data using information on marital satisfaction gathered from 218 continuously-married couples in the Early Years of Marriage Project (EYMP).

Traditional Random- and Fixed-Effects Models for Longitudinal Data

In much of the family literature the most common procedure for examining repeated observations on individuals or other units of observation is a REM or a FEM. Many statistical packages easily allow estimation of REMs and FEMs. STATA (XTREG)

J. Teachman (✉)
Department of Sociology, Western Washington University, Bellingham, WA, USA
e-mail: Jay.Teachman@wwu.edu

S. M. McHale et al. (eds.), *Emerging Methods in Family Research,*
National Symposium on Family Issues 4, DOI 10.1007/978-3-319-01562-0_1,

and SAS (PROC GLM for fixed effects and PROC MIXED for random effects) are popular options. It is also possible to estimate REMs and FEMs using structural equation models (Bollen and Brand 2010; Teachman et al. 2001) using programs such as Mplus, EQS, AMOS, or PROC CALIS in SAS. SEMs allow researchers to explicitly model or manipulate co-variances, and as I demonstrate, SEMs allow hybrid models mixing both fixed and random effects, as well as extensions to models such as LGCMs. Accordingly, in this chapter all models are presented and estimated as SEMs.

To begin the discussion and fix ideas, consider the following REM:

$$y_{it} = \alpha_t + \beta_{\mathbf{yxt}}\mathbf{X_{it}} + \beta_{\mathbf{yzt}}\mathbf{Z_i} + \beta_{y\eta t}\eta_i + \varepsilon_{it} \tag{1.1}$$

y_{it} is the value of the dependent variable for the i^{th} case at time t; α_t is an intercept term at time t; $\mathbf{X_{it}}$ is a vector of time-varying covariates for the i^{th} case at time t; $\beta_{\mathbf{yxt}}$ is a vector of coefficients indicating the effects of $\mathbf{X_{it}}$ on y_{it}; $\mathbf{Z_i}$ is a vector of time-constant covariates for the i^{th} case; $\beta_{\mathbf{yzt}}$ is a vector of coefficients indicating the effects of $\mathbf{Z_i}$ on y_{it}; η_i is a scalar indicating all of the latent time-constant factors affecting y_{it}; $\beta_{y\eta t}$ is the coefficients linking the latent factor η_i to y_{it} at time t (here all values of this vector are set equal to 1.0); and ε_{it} is a random disturbance for the i^{th} case at time t with $E(\varepsilon_{it}) = 0$ and $E({\varepsilon^2}_{it}) = {\sigma^2}_{\varepsilon t}$. It is assumed that ε_{it} is uncorrelated with $\mathbf{X_{it}}$, $\mathbf{Z_i}$, and η_i, that $COV(\varepsilon_{it},\varepsilon_{it}) = 0$ for $t \neq s$, and that η_i is uncorrelated with $\mathbf{X_{it}}$ and $\mathbf{Z_i}$. This is the default REM estimated by most software products in which the effects of the time-varying variables are constrained to be constant across time, as are the variances of the error terms. This is also the REM that many, if not most, researchers using longitudinal data report. The key assumption for the purposes of this chapter is that η_i is uncorrelated with the included covariates. Accepting this assumption means that the standard errors of the coefficients are adjusted for clustering but are not adjusted for unmeasured covariates that may be correlated with both the dependent variable and the covariates.

A FEM can be written as:

$$y_{it} = \alpha_t + \beta_{\mathbf{yx}}\mathbf{X_{it}} + \beta_{y\eta}\eta_i + \varepsilon_{it} \tag{1.2}$$

The $\mathbf{Z_i}$ are now dropped from the equation because they are assumed to be included in the time-constant latent variable η_i. The key assumption here is that the model allows η_i to be correlated with the $\mathbf{X_{it}}$. Note that even though one loses the ability to obtain an estimate of the impact of specific time-constant variables on the outcome, their effects are still controlled by including η_i in the model. In a FEM, the effects of the time-varying variables are washed of any effect linked to unmeasured time-constant factors. Without prior evidence, I suggest that assuming zero covariance between latent time-constant factors and the included covariates is risky. If a correlation exists, and it is not modeled, the estimated parameter estimates will be biased, either upward or downward. The FEM avoids this issue.

If the REM does not include $\mathbf{Z_i}$ then it is simply a restricted version of the FEM that constrains the covariances between η_i and $\mathbf{X_{it}}$ to be zero. In other words, the models are nested and can be compared via a standard likelihood ratio test. If the

Table 1.1 Results from estimating random- and fixed-effects models for BMI

	Random effects	Fixed effects
Cohabiting	0.264**	0.280**
Married	0.485**	0.475**
Highest grade completed	−0.112**	0.087
T1 intercept	27.847**	25.171**
T2 intercept	28.729**	26.032**
T3 intercept	29.566**	26.855**
T4 intercept	29.894**	27.175**
LR chi-square	682.13	651.58
Df	53	41
RMSEA	0.082	0.092
BIC	286.03	345.16

** $p < 0.05$

REM includes $\mathbf{Z_i}$ but the FEM does not, then the models are not nested. On the other hand, if the REM includes $\mathbf{Z_i}$ and the FEM also includes $\mathbf{Z_i}$ (by constraining the covariance between $\mathbf{Z_i}$ and η_i to equal zero) the two models are once again nested (the result is a hybrid REM/FEM). Bollen and Brand (2010) provide an overview of these points within the context of a general model for panel data.

Results from estimating these two models using the NLSY data on BMI are shown in Table 1.1. The NLSY data consist of 1761 observations for men in 1992, 1996, 2000, and 2004. Cases with missing data were deleted, as were men with BMI values greater than 50. Marital status is time-varying and is measured as married, cohabiting, and other. Highest grade completed is time varying and indicates the highest year of schooling completed by the respondent. A time-constant indicator of race/ethnicity measured as Black, Hispanic, and other is included. PROC CALIS in SAS was used to generate these estimates. The estimates are identical to models estimated using XTREG in STATA. According to the REM, when compared to men not in a union, both cohabitors and married men are heavier, and men with more education have lower values of BMI. The FEM shows similar estimates of the effects of marital status, but the effect of highest grade completed has changed signs and is no longer statistically significant, indicating that its effect can be attributed to the common latent factor. The intercept terms indicate that for both the REM and the FEM there is a tendency for BMI to increase over time. A variety of fit statistics are provided for each model: LR chi-square, Root Mean Square of Approximation (RMSEA), and Bayesian Information Criterion (BIC). Differences between the LR chi-square values suggest that the FEM is a better choice than the REM ($X = 30.55$, 12 df). The BIC value for the FEM is much larger than the BIC value for the REM however, indicating some ambiguity in whether the FEM should be favored. In large part, this is a common occurrence because FEMs use additional degrees of freedom (by allowing non-zero covariances between the latent term and the time-varying covariates), and BIC penalizes models that use more degrees of freedom. Overall, the fit statistics do not indicate well-fitting models though. In particular, BIC values should be negative for models that fit the data well.

Latent Growth Curve Models for Longitudinal Data

The lack of fit for either the REM or FEM in Table 1.1 suggests that another model specification is in order. I argue that in the case of a variable like BMI a LGCM is appropriate. LGCMs are appropriate when the outcome variable being considered follows a trajectory of change across time (i.e., does not randomly shift across time). A simple LGCM can be expressed as follows (ignoring time-constant variables):

$$y_{it} = \beta_{\mathbf{yxt}}\mathbf{X_{it}} + \beta_{y\eta 0t}\eta_{oi} + \beta_{y\eta 1t}\eta_{1i} + \varepsilon_{it} \tag{1.3}$$

η_{oi} is a latent factor indicating initial values of BMI with slopes, $\beta_{y\eta 0t}$, constrained to equal 1, and η_{1i} is a second latent factor with slopes, $\beta_{y\eta 1t}$, indicating change in BMI over time. All other terms and assumptions are defined as in Eq. 1.1 with the caveat that the latent factor is now represented by two terms. Most researchers call η_{oi} the intercept (beginning or initial value of the outcome) and η_{1i} the slope of the model (change across time from the initial value of the outcome). In a LGCM therefore, there are two latent components rather than one as is the case in a REM. Similar to a REM, one latent factor (intercept) describes a stable component across time. (Slopes are constrained to unity.) The second latent factor allows variation from this stable component over time and can be thought of as representing the rate of change across time. This is the factor that models structured change across time.

LGCMs (as well as REMs and FEMs) are hierarchical linear models (HLM) in that there are two levels of variation represented: within-subject and between-subject. The $\mathbf{X_{it}}$ represent within-subject variation, whereas η_{oi} and η_{1i} represent between-subject variation. Because they vary between subjects, both η_{oi} and η_{1i} can be represented as functions of other time-constant covariates. This point is demonstrated later in this chapter. Many applications of REMs and FEMs ignore the fact that variation in outcomes occurs both within and between respondents.

A graphic representation of this model is shown in Fig. 1.1. Note that as shown this is a random-effects LGCM because both latent terms are assumed to be independent of any covariates. Also note that the slopes for the second latent term representing change in BMI are fixed at 0, 1, 2, and 3 to reflect a linear trajectory of gains in BMI. (Alternative specifications are possible.) The model shown is known as a conditional LGCM because the latent terms are estimated conditional on the effects of marital status and education (and vice versa). If marital status and education were not included in the model and each BMI included an error term instead, the model would be an unconditional LGCM. For the purpose of establishing a baseline, I estimate an unconditional LGCM using the NLSY data on BMI. The resulting value for η_{oi} is 26.80 and the resulting value for η_{1i} is 0.693. These values can be thought of as the average value of the intercept and slope, respectively. That is, if all 1761 cases were plotted, the average starting value for BMI would be 26.80 and the average slope indicating gain in BMI over time would be 0.693. Of course, these average values also have standard errors because they are not fixed across individuals. In this case, the standard error for the intercept term is 0.099, and the standard error for the slope term is 0.023. In both cases, a simple t-test indicates that there is statistically

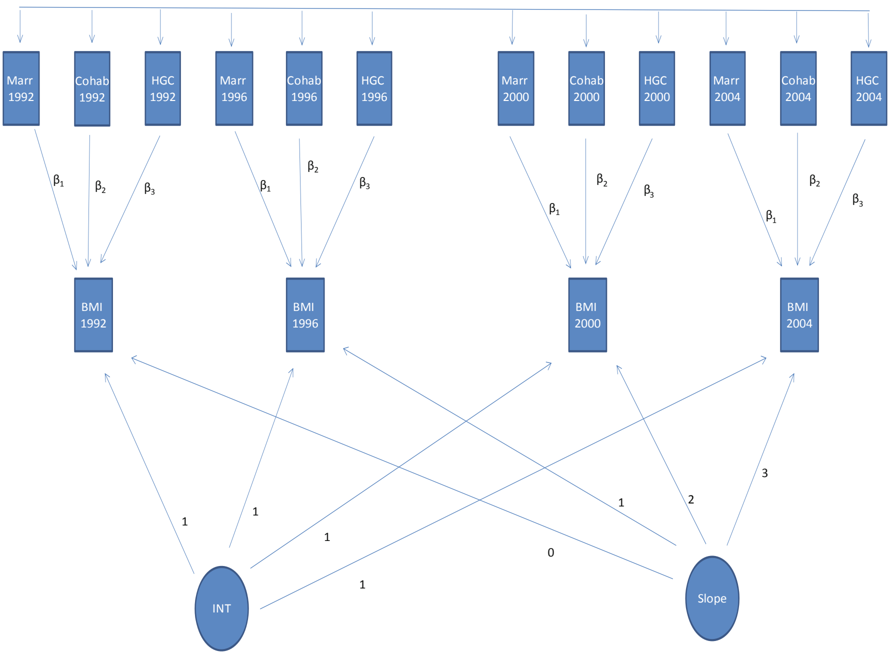


Fig. 1.1 Simple latent growth curve model with marital status and highest grade of schooling completed as time-varying covariates

significant variation in both initial value of BMI and rate of change across time (slope). The model fit statistics for the unconditional model are 182.64/8 for the LR Chi-square, 0.111 for RMSEA, and 122.85 for BIC.

The fit statistics for the unconditional LGCM are not particularly good and suggest that the model can be improved. Accordingly, I estimate the conditional LGCM represented in Fig. 1.1. The model fit statistics for this model are shown in Table 1.2 for Model A. These values indicate that this model is a better fit to the data. In particular, RMSEA is much lower (0.045 vs. 0.111), and BIC is now negative (−153.75). The parameter estimates for this model are shown in Table 1.3. The latent intercept has a value of 28.044 and the slope estimate is 0.697. Thus, the addition of the time-varying covariates does not dramatically alter estimates of these basic parameters. Yet, the fact that Model A fits the data better than an unconditional model indicates that net of the latent growth factors, marital status and education significantly affect BMI. Compared to the coefficients for traditional REM shown in Table 1.1, the effects of being married (0.515) and the effects of education (−0.118) are similar; however, the effect of cohabitation (0.135) is much smaller and not statistically significant. This latter result suggests that the effect of cohabitation as estimated in the traditional REM can be attributed to latent growth in BMI over time. That is, beyond the tendency for men to become heavier as they age, cohabitation does not appear to be related to BMI.

Table 1.2 Model fit statistics for various latent-growth curve models: NLSY data on male BMI

Model	LR chi-square/df	RMSEA	BIC
A. LGC REM	242.36/53	0.045	−153.75
B. LGC REM + quadratic	102.77/49	0.025	−263.44
C. LGC FEM	198.58/29	0.058	−18.16
D. LGC FEM + quadratic	61.98/25	0.020	−124.87
E. hybrid LGC + quadratic	83.63/45	0.022	−252.68
F. hybrid LGC + quadratic + mediated race/ethnicity	87.11/49	0.021	−279.10
G. hybrid LGC + quadratic + direct race/ethnicity equal slopes	101.02/51	0.024	−280.13
H. hybrid LGC + quadratic + direct race/ethnicity unequal slopes	72.18/45	0.019	−264.13
I. hybrid LGC + quadratic + direct race/ethnicity unequal slopes + unequal variances	38.59/42	0.000	−275.30
J. hybrid LGC + quadratic + direct race/ethnicity unequal slopes + unequal variances + level 1 unequal slopes	30.96/33	0.000	−215.67
K. hybrid LGC + quadratic + mediated race/ethnicity + unequal variances	53.81/46	0.010	−289.98

Some researchers might stop here after concluding that marriage and education affect BMI but that cohabitation does not. However, there are important extensions to Model A. One possible extension is to consider non-linear changes in BMI over time. There are many ways to allow for non-linear change in the outcome variable but perhaps the most parsimonious is to model a quadratic rate of change (by adding a quadratic latent term to Eq. 1.3). In this case, the additional latent construct allows for a quadratic change in the slope, in which each slope is just the square of the linear change in slope. (i.e., Each of the paths from the latent quadratic slope construct is the square of the corresponding latent linear slope construct). As shown in Table 1.2, this model (Model B) fits the data better than a model with only a linear term. RMSEA is now 0.025 (vs.0.045), and the value of BIC is more negative (−263.44 vs. −153.75). Table 1.3 shows the parameter estimates for this model. The intercept term is 28.171, the linear slope is 1.115, and the quadratic slope is −0.139, all statistically significant. The positive linear slope and negative quadratic slope indicate that BMI tends to increase over time but at a diminishing rate. The coefficients for marital status and education remain similar to those estimated for Model A, and the effect of cohabitation remains non-significant.

A Latent Growth Curve Fixed-Effects Model for Longitudinal Data

Another extension of the LGC REM model is to consider a LGC FEM model. As indicated in Fig. 1.1, in a LGC REM there are no covariances allowed between the latent terms and any of the time-varying covariates. If these covariances are allowed

Table 1.3 Estimated components of various latent-growth curve models: NLSY data on male BMI

Model components	Model A	Model B	Model C	Model D	Model E	Model F	Model G	Model H	Model I	Model J	Model K
Manifest variable equations:											
Cohabiting	0.135	0.148	0.010	0.024	0.175	0.176**	0.164	0.187*	0.187*		0.176*
Time 1										0.367**	
Time 2										0.198	
Time 3										0.121	
Time 4										0.242	
Married	0.515**	0.490**	0.533**	0.483**	0.483**	0.499**	0.495**	0.498**	0.493**		0.494**
Time 1										0.602**	
Time 2										0.502**	
Time 3										0.434**	
Time 4										0.383**	
HGC	−0.118**	−0.137**	0.067	0.003	0.058	0.078	0.078	0.078	0.068		0.063
Time 1										0.061	
Time 2										0.081	
Time 3										0.082	
Time 4										0.050	
Black							1.010**				
Time 1								0.896	0.885**	0.883**	
Time 2								0.903**	0.900**	0.932**	
Time 3								1.373**	1.370**	1.339**	
Time 4								1.461**	1.451**	1.441**	
Hispanic							1.569**				
Time 1								1.563**	1.547**	1.530**	
Time2								1.450**	1.441**	1.480**	
Time 3								1.787**	1.779**	1.767**	
Time 4								1.678**	1.663**	1.659**	
Latent Variable Equations:											
Intercept	28.044**	28.171**	25.585**	26.376**	25.545**	24.838**	24.800**	24.836**	24.984**	24.994**	24.987**
Black						0.850**					0.857**

Table 1.3 (continued)

Model components	Model A	Model B	Model C	Model D	Model E	Model F	Model G	Model H	Model I	Model J	Model K
Hispanic						1.519**					1.520**
Slope	0.697**	1.115**	0.682**	1.097**	1.091**	1.032**	1.088**	1.045**	1.035**	0.637*	1.009**
Black						0.215**					0.213**
Hispanic						0.068					0.068
Quadratic slope		−0.139**		−0.136**	−0.136**	−0.135**	−0.135**	−0.140**	−0.136**	−0.031	−0.128**
Error BMI1	2.024**	1.703**	2.021**	1.702**	1.703**	1.703**	1.702**	1.689**	0.890*	0.889	0.898*
Error BMI2	2.024**	1.703**	2.021**	1.702**	1.703**	1.703**	1.702**	1.689**	1.536**	1.535**	1.551**
Error BMI3	2.024**	1.703**	2.021**	1.702**	1.703**	1.703**	1.702**	1.689**	1.870**	1.868**	1.883**
Error BMI4	2.024**	1.703**	2.021**	1.702**	1.703**	1.703**	1.702**	1.689**	2.228**	2.239**	2.245**
Error intercept	15.493**	14.971**	15.843**	15.178**	15.364**	15.070**	15.073**	15.084**	15.782**	15.768	15.772**
Error slope	0.527**	1.671**	0.526**	1.660**	1.656**	1.652**	1.656**	1.687**	2.391**	2.377*	2.363**
Error quadratic slope		0.141**		0.141**	0.141**	0.141**	0.141**	0.145**	0.160**	0.157**	0.156**
Covariances:											
Intercept—slope	0.097	0.758**	−0.058	0.758**	0.763**	0.737**	0.733**	−0.015	−0.024	−0.029	−0.015
Intercept—quadratic slope		−0.297**		−0.295**	−0.300**	−0.299**	−0.300**	−0.128	−0.125	−0.122	−0.128
Slope—quadratic slope		−0.392**		−0.389**	−0.390**	−0.391**	−0.890**	−0.527**	−0.537**	−0.532**	−0.527**
Intercept—time 1 HGC					−1.582**	−1.470**	−1.469**	−1.470**	−1.422**	−1.388*	−1.421**
Intercept—time 2 HGC					−1.528**	−1.420**	−1.420**	−1.420**	−1.370**	−1.335**	−1.369**
Intercept—time 3 HGC					−1.551**	−1.449**	−1.448**	−1.449**	−1.398**	−1.363**	−1.398**
Intercept—time 4 HGC					−1.592**	−1.487**	−1.487**	−1.487	−1.438	−1.403	−1.438**

* $p < 0.10$
** $p < 0.05$

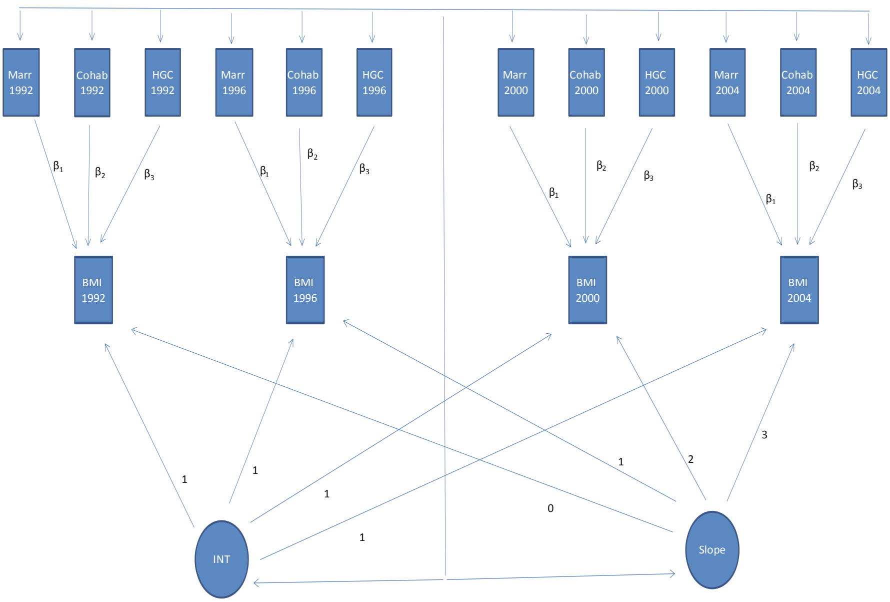


Fig. 1.2 Fixed-effects latent growth curve

to differ from zero then a LGC FEM model results. As far as I am aware, such a model has not been presented in the existing literature dealing with LGCMs even though the extension from an LGC REM is straightforward. A LGC FEM is illustrated in Fig. 1.2 where non-zero covariances between the latent intercept and slope terms and the time-varying covariates are allowed. The value in estimating a LGC FEM is that it will provide unbiased estimates of the effects of included covariates if they are correlated with unmeasured latent terms.

The fit statistics for the LGC FEM shown in Fig. 1.2 are presented in Table 1.2 (Model C). Compared to Model A, the RMSEA is larger (0.058 vs. 0.025) and BIC is less negative (-18.16 vs. -263.44). The parameter estimates for this model shown in Table 1.3 indicate an intercept term of 25.585 and a slope of 0.682, values similar to the LGC REM. The effects of cohabitation and marriage are also similar to those estimated in the LGC REM (although the non-significant effect of cohabitation is now larger), but the effect of education is no longer negative and statistically significant, indicating that its impact was due to covariation with the latent factor.

An improvement in model fit results when estimating a LGC FEM that allows a latent quadratic slope term. For the sake of parsimony, when estimating this model I allowed only the latent intercept and latent linear slope constructs to be correlated with the time-varying covariates. The fit statistics for Model D in Table 1.2 (LR chi-square 61.98/25, RMSEA $= 0.020$, and BIC $= -124.87$) indicate a better fit to the data than Model C. The parameter estimates shown in Table 1.3 indicate that, similar to Model C, the only time-varying covariate to affect BMI is being married. Although the LGC FEM in Model D with a quadratic term yields a better fit than the LGC FEM in Model C with only a linear term, the fit compared to the LGC REM with a quadratic term (Model B) is equivocal. Whereas the RMSEA is smaller (0.020 vs. 0.025), the difference in BIC statistics is considerable with the REM version possessing a much more negative BIC value (−263.44 vs. −124.87). The difference in BIC values suggests an elaboration of the LGC FEM that may lead to a better fitting model.

Specifically, the LGC FEM estimated in Model D includes a sizeable number of estimated covariances between the latent constructs and the time-varying covariates (even though the latent quadratic term was not allowed to covary with these covariates). As indicated earlier, BIC strongly penalizes models that estimate more parameters than necessary. If some of the covariances are statistically indistinguishable from zero, then valuable degrees of freedom are being wasted. When I examined the covariances estimated in Model D, I found that most of them were not distinguishable from zero. Indeed, the only covariances that were consistently significant involved schooling and the latent intercept term. Fit statistics for a LGC FEM allowing only these covariances between latent and observed terms are shown in Model E of Table 1.2. I term this a hybrid LGC because it involves only a subset of all possible covariances between the latent and observed variables. Compared to Model B, the LR chi-square value (83.63/45) and the RMSEA value (0.022) indicate a better fit to the data. The BIC value for this model (−252.68) is not quite as negative as the BIC value for the LGC REM with a quadratic term (−263.44) but is a significant improvement over earlier versions of the LGC FEM.

Parameter estimates for Model E are shown in Table 1.3. Also shown in Table 1.3 for Model E are the covariances between education and the latent intercept term. Each of the four covariances is statistically significant and negative. In other words, there are unmeasured factors that link having more education with lower body weight.

Model E suggests two important points with respect to the effects of the time-varying covariates. First, the effect of marital status is not substantially biased by failure to include covariances with the latent terms in the model. Thus, we can be more confident in our ability to state that being married is positively linked to BMI whereas cohabitation is not. Second, the effect of education is biased by the failure to include covariances with the latent terms in the model (here the latent intercept term).

Time-Constant Covariates in Latent Growth Curve Models with Fixed Effects

An extension often found in standard LGC REMs is to allow the latent constructs to be functions of time-constant observed variables. In other words, between respondent variation in the outcome under consideration can be modeled. Consider the following equations:

$$y_{it} = \beta_{\mathbf{yxt}} X_{it} + \beta_{\mathbf{y}\eta\mathbf{0t}} \eta_{oi} + \beta_{\mathbf{y}\eta\mathbf{1t}} \eta_{1i} + \varepsilon_{it} \tag{1.4}$$

$$\eta_{oi} = \alpha_{00} + \gamma_{\eta 0z} Z_i + \delta_{0i} \tag{1.5}$$

$$\eta_{1i} = \alpha_{10} + \gamma_{\eta 1z} Z_i + \delta_{1i} \quad (1.6)$$

α_{00} and α_{10} are constant terms; $\gamma_{\eta 0z}$ and $\gamma_{\eta 1z}$ are slopes; Z_i is a time-constant variable affecting the latent intercept and slope; δ_{0i} and δ_{0i} are error terms; and all other terms are as defined earlier. Equations 1.4–1.6 emphasize the hierarchical nature of the LGC framework, which uses information both within- and between-subjects. Within-subject variation is modeled as a function of the latent intercept and slope, as well as time-varying covariates. Between-subject variation in the latent intercepts and slopes is modeled as a function of variation on time-constant variables such as race or ethnicity.

Assuming that Z_i is a measure of race/ethnicity (Black and Hispanic), this model assumes that the effects of race/ethnicity on BMI are mediated by the latent intercept and slopes. Fit statistics for this model are presented in Table 1.2 (Model F). Compared to previous models, this model fits the data well with LR chi-square = 87.11/49, RMSE = 0 .021 and BIC = −279.10. Parameter estimates for this Model F are shown in Table 1.3. The results indicate that Blacks (0.850) and Hispanics (1.519) have higher initial levels of BMI, with the value for Hispanics being nearly twice as high as that for Blacks. The pace of increase in BMI is greater for Blacks (0.215) compared to Whites but does not differ for Hispanics. For the sake of simplicity, I did not allow race/ethnicity to affect the quadratic slope term.

Although not common in the literature, it is possible to allow time-constant variables to affect the time-varying dependent variable(s) directly. No special accommodations are necessary to do this in the LGC REM. In the LGC FEM, however, covariances between the latent terms and the time-constant variable must be set to zero in order for the model to be identified. The mediated model assumes that all of the effects of the time-constant variable are captured by its impact on the latent growth parameters. This assumption is strong and it may well be the case that the time-constant variable directly affects the outcome variables being examined and not the intercept and trajectory of change.

A model that includes both the direct and mediated effects of a time-constant variable is not identified. Thus, researchers will need to choose between the two. Fortunately, it has been shown that the mediated model is nested within the direct model (Stoel et al. 2004) This nesting means that a LR chi-square test can be used to determine whether the direct model is warranted. Table 1.2 presents fit statistics for a model allowing direct effects of race/ethnicity on BMI (Model G). This model constrains the effect of race/ethnicity to be constant across all four measurement points. The LR chi-square value (101.02/51) and the RMSEA (0.024) indicate that the mediated model should be preferred. (There is little difference between BIC values.) The parameter estimates for Model G in Table 1.3 indicate that the effect of being Black (1.010) or Hispanic (1.569) is to increase BMI at each point in time. Estimates of the latent growth parameters are similar to previous estimates.

As noted, Model G constrains the direct effects of race/ethnicity to be equal across all four time points. It may be plausible that these effects differ across time. Accordingly, Model H represents a model where the effects of race/ethnicity are free to vary over time. This model fits the data better than the mediated model according to the LR chi-square (72.18/45 vs. 87.11/49) and RMSEA (0.019 vs. 0.021). Because more degrees of freedom are being used though, the BIC value (−264.13 vs. −279.10) indicates a preference for the mediated model. Additional elaborations are possible. Model I in Table 1.2 extends Model H by allowing unequal variances for the error terms associated with each value of BMI. Model J continues relaxing assumptions in that the time-varying predictors are allowed to have effects on BMI that vary across time. Model fit statistics indicate that these models do not yield an improvement in fit to the data, and I do not discuss them further.

A final model (K) is shown in Table 1.2. In this model, the effects of race/ethnicity are mediated through the latent terms, and unequal error variances in BMI are allowed. The LR chi-square value (53.81/46) and RMSEA value (0.010) when compared to Model I do not indicate a better fit to the data. The value of BIC (− 289.98), however, indicates that this model may be the preferred option. The parameter estimates shown in Table 1.3 indicate parameter estimates very similar to those shown for Model F. Hispanics have the highest initial levels of BMI followed by Blacks and then Whites. Compared to Whites the rate of change in BMI is steeper for Blacks but not Hispanics. The freed variances for the four measures of BMI continue to indicate increasing random variation over time. In terms of marital status and education, being married or in a cohabiting relationship increases BMI, whereas there is no effect of schooling once a fixed-effects estimator is employed.

A Latent Growth Curve Model for Paired Data

I have outlined what I believe to be some valuable extensions of LGCMs for family data, using data on BMI and marital status taken from the National Longitudinal Study of Youth (NLSY). I now offer a further extension with a simple example. The extension I propose is for paired or matched family data. That is, data that refer to two or more members of the same family. Neale and McArdle (2000) have shown how LGCMs can be used to examine twin data. An empirical example of this procedure is provided by Hjelmborg et al. (2008) who demonstrate that genetic influences on rate of change in BMI are different from those affecting level of BMI. In essence this procedure takes advantage of the fact that multilevel models, including LGCMs, can be estimated simultaneously for multiple groups. A model is estimated simultaneously for several groups (e.g., identical twins, fraternal twins) and constraints are imposed on the various parameters of the model across groups in order to determine whether the models for particular twin groups differ from those of other twin groups.

Although this is a very useful approach and one that demonstrates the ability of the model to be estimated across groups, it does have limitations. In particular, such a model does not allow the parameters of one group to affect the parameters of another group, much like one would anticipate in a family group. Similarities are assumed to be a function of shared genetic potential rather than patterns of interaction. This limitation may make sense when examining twin data but is less useful in other circumstances. For example, consider the case of a married couple

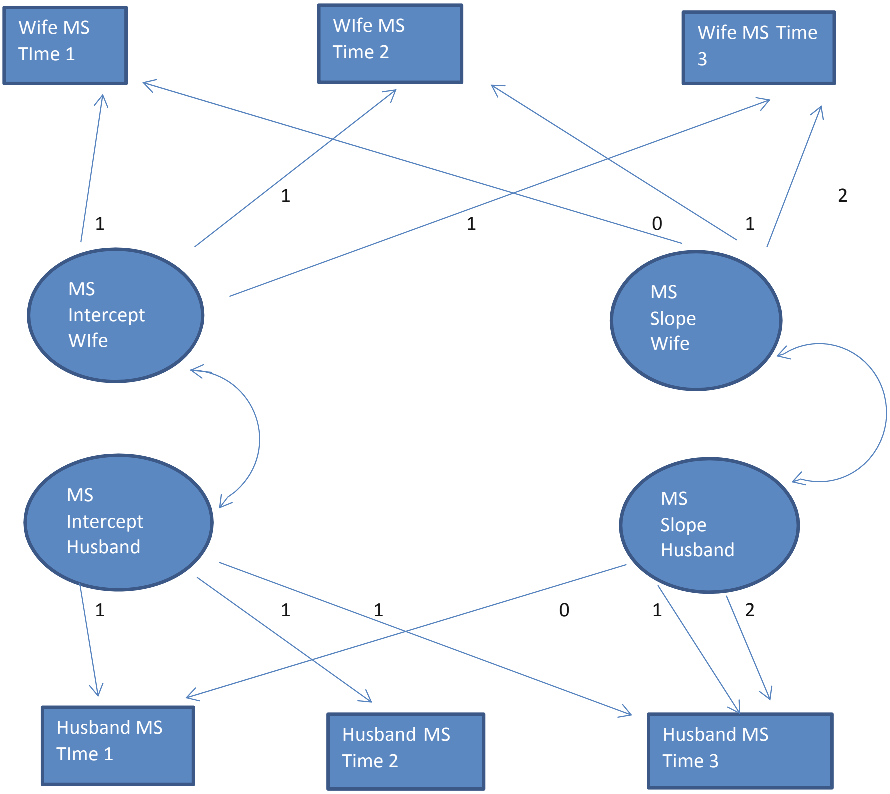


Fig. 1.3 Latent growth curve model for paired processes

and their respective paths of change in marital satisfaction over time. It may be the case that the trajectory of marital satisfaction in one spouse affects or is affected by the trajectory of the other spouse.

A generic example of this notion is illustrated in Fig. 1.3. Walker et al. (1996) provide an example of this sort of model using data on two variables (amount of caregiving performed and satisfaction with caregiving) measured for a single individual. In Fig. 1.3, I assume there is longitudinal information obtained on one variable (here marital satisfaction) for two related individuals (husband and wife). A very simple model is presented with no time varying covariates affecting the measured variable of interest. It is a simple extension of Fig. 1.3 to include time varying covariates and thus have the ability to estimate REM or FEM LGCMs in this framework as described earlier.

The most important components of the model in Fig. 1.3 for the purposes of matched processes are the relationships allowed between the intercepts and slopes across the two persons in the model. Two-way arrows are shown indicating no assumed directionality of effects (i.e., simply the covariance between the two terms).

With proper theoretical justification and appropriate specification, effects in each direction could be estimated. The arrow linking the intercepts indicates matching on the underlying value of the variable in question. Using the marital satisfaction of married couples, the covariance between intercepts can be interpreted to illustrate the degree of marital matching on relationship quality. The covariance between the slopes can be interpreted to illustrate the extent to which the paths of change in marital satisfaction within a married couple are linked.

I estimated such a model using four waves of data taken from the Early Years of Marriage Project (EYMP), (Veroff et al. 1986–1989) Data on marital satisfaction were collected in 1986, 1987, 1988, and 1989. Using information provided by 218 continuously-married couples, I computed a simple additive marital satisfaction scale based on four items: (1) how likely do you believe your marriage will be intact in five years, (2) how stable do you perceive your marriage to be, (3) have you ever considered leaving your marriage, and (4) how satisfied are you with your marriage. Higher scores indicate a greater degree of marital *dissatisfaction*.

The model estimated corresponds to that shown in Fig. 1.3 using four time points in the EYMP. I did not attempt to find a best fitting model, but I did compare two models that were otherwise identical. The first model constrained the covariances between the intercepts and slopes of husbands and wives to be equal to zero. The second model allowed these covariances to vary. The second model fitted the data much better ($X^2 = 120.35$ with 2 df). The covariance between the intercepts was 0.60, and the covariance between the slopes was 0.16. The slopes were positive (1.05 for wives and 0.87 for husbands) reflecting growing marital dissatisfaction over time. The positive correlation between the slopes for the spouses indicates that the pace at which one spouse's marital dissatisfaction grows influences the pace at which the other spouse's dissatisfaction grows. Failure to consider the joint influence of each spouse on the other would yield a biased estimate of the pace of change in marital dissatisfaction. (Here, failure to include the positive covariance yields an overestimate of the trajectory of change in marital dissatisfaction for each spouse—results not shown.) More importantly, failure to include the covariance yields an unrealistic model where the marital satisfaction of each spouse supposedly unfolds over time in a manner not affected by the other spouse's marital satisfaction.

Discussion

In this chapter, I have outlined the straightforward extension of REMs and FEMs to LGCMs. Whereas, LGCMs are prevalent in some disciplines such as developmental psychology, they remain rare in family studies. If the phenomenon under study changes across time in a systemic fashion, LGCMs offer the opportunity to model this change. Existing versions of the LGCM are all in the REM framework. That is, the latent terms describing initial levels and change in levels of the dependent variable are assumed to be independent of any other covariates affecting the dependent variable. This is a strong assumption and violations may lead to biased estimates of parameters.

I show that the SEM framework provides researchers a way to test this assumption by allowing estimation of fixed-effect versions of the LGCM. By allowing all or some of the time-varying covariates to covary with the included latent terms, LGC FEM provides a powerful tool, allowing researchers to account for patterns of covariation that may bias parameter estimates.

Further, I show how time-constant covariates can be included in LGCMs. The effects of time-constant covariates may be expressed directly on the latent parameters of the model or directly on the outcome under consideration. Because these two models are nested, a simple LR Chi-square test can be used to adjudicate between the two.

Finally, I show how matched or paired data can be examined within the framework of LGCMs. These models assume that the parameters underlying the trajectory of change across time for one partner affect the same parameters for the other partner. Failure to account for the covariation of these parameters can lead to biased estimated of the underlying parameters.

References

Allison, P. (2005). *Fixed effects regression models for longitudinal data using SAS*. Cary: SAS Institute Press.

Bollen, K. A., & Brand, J. E. (2010). A general panel model with random and fixed effects: A structural equations approach. *Social Forces, 89,* 1–34.

Hjelmborg, J., Fagnani, C., Silventoinen, K., McGue, M., Korkeila, M., Christensen, K., Rissanen, A., & Kaprio, J. (2008). Genetic influences on growth traits of BMI: A longitudinal study of adult twins. *Obesity, 16,* 847–852.

Lyons, K., & Sayer, A. (2005). Longitudinal dyad models in family research. *Journal of Marriage and Family, 67,* 1048–1060.

Neale, M., & McArdle, J. (2000). Structural latent growth curves for twin data. *Twin Research, 3,* 165–177.

Stoel, R. D., van den Wittenboer, G., & Hox, J. (2004). Including time-invariant covariates in the latent growth curve model. *Structural Equation Modeling, 11*(2), 155–167.

Teachman, J., Duncan, G., Yeung, W. J., & Levy, D. (2001). Covariance structure models for fixed and random effects. *Sociological Methods and Research, 30,* 271–288.

Umberson, D., Liu, H., & Powers, D. (2009). Marital status, marital transitions, and body weight. *Journal of Health and Social Behavior, 50,* 327–343.

Veroff, J., Douvan, E., Hatchett, S., & Orbuch, T. (1986–1989). Early Years Of Marriage (Eym) Project (years 1–4). Ann Arbor, MI: University of Michigan, Institute for Social Research, Survey Research Center [producer], 2007. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2007-10-03. doi:10.3886/ICPSR04557.v1.

Walker, A., Acock, A., Bowman, S., & Li, F. (1996). Amount of care given and caregiving satisfaction: A latent growth curve analysis. *Journal of Gerontology: Psychological Sciences, 51B*, 130–142.

Chapter 2
Families as Coordinated Symbiotic Systems: Making use of Nonlinear Dynamic Models

Nilam Ram, Mariya Shiyko, Erika S. Lunkenheimer, Shawna Doerksen and David Conroy

Families as Coordinated Symbiotic Systems: Making use of Nonlinear Dynamic Models

Families are often conceptualized as continually evolving, relational systems (Cox and Paley 1997; Minuchin 1985). Individual members influence and are influenced by all other members. These reciprocal relations coalesce into family-level symbiotic processes and are the core of study in family systems research and therapy (see Lunkenheimer et al. 2012). Wohlwill (1991) noted that, "... what [reciprocal relationships] would call for are methodologies that allow one to model the interpatterning between two [or more] sets of processes each of which is undergoing change, in part as a function of the other ... The closest approach to this kind of modeling that is indicated for this purpose are probably some of the models from the field of

N. Ram (✉)
Department of Human Development and Family Studies, The Pennsylvania State University, University Park, PA, USA
e-mail: nur5@psu.edu

M. Shiyko
Department of Applied Counseling and Educational Psychology, Northeastern University, Boston, MA, USA
e-mail: M.Shiyko@neu.edu

E. S. Lunkenheimer
Department of Human Development & Family Studies, Colorado State University, Fort Collins, CO, USA
e-mail: erika.lunkenheimer@colostate.edu

S. Doerksen
Department of Recreation, Park and Tourism Management, The Pennsylvania State University, University Park, PA, USA
e-mail: sxd49@psu.edu

D. Conroy
Department of Kinesiology, The Pennsylvania State University, University Park, PA, USA
e-mail: dec9@psu.edu

S. M. McHale et al. (eds.), *Emerging Methods in Family Research*, National Symposium on Family Issues 4, DOI 10.1007/978-3-319-01562-0_2,

ecology and similar systems-analytical work" (p. 128). Following this suggestion we began exploring how the theoretical and analytical approaches used in ecology relate to study of person-context transactions and longitudinal structural equation models (Ram and Nesselroade 2007; Ram and Pedersen 2008). Here we extend our thinking to the modeling of family systems.

Theory: Families as Dynamic Systems

Over the last half-century, developmental scientists have often argued for and sometimes formulated dynamic systems and process-based accounts of development (Ford and Lerner 1992; Gottlieb 1997; Sameroff 1983; Schneirla 1957; Werner 1957). Likewise, family therapists and researchers often promote a systems view, stressing that the family is a whole, has interdependent subsystems, and that interactions among these subsystems are governed by homeostatic and circular processes that emerge and adapt to internal and external demands (e.g., Minuchin 1974, 1985). Families are dynamic systems capable of both (a) adaptive self-stabilization, wherein coordinated action of subsystems compensate for changing conditions in the environment and maintain homeostasis, and (b) adaptive self-organization, wherein the system as a whole transforms to accommodate change in or challenge to the existing configuration (see Cox and Paley 1997, Thelen and Ulrich 1991).

These principles provide a framework for examining change within family systems at multiple time scales (Granic 2005; van Geert 1998). Shorter-term, micro-level changes indicate how a family system self-stabilizes and functions as a unit to achieve homeostasis. For instance, consider a couple who tells jokes every evening as a way to lift each others' mood and decompress from the day's stress. When a major change occurs, the system may self-organize into a qualitatively different pattern of self-stabilization. When a new child joins the family, the sharing of jokes is replaced by a new homeostasis–diaper changing and fatigue. If we explicitly consider the dynamic and reciprocal interactions among family members as the actions of a dynamic system that self-stabilizes and self-organizes, we can begin to model the structure, organization, and patterning of these relationships within and across time and context (Fogel 1993; Lunkenheimer and Dishion 2009).

Methods: Nonlinear Dynamic Models

Applications to Biological Systems

As does family systems theory, biological theory underscores the importance of dynamic interactions. However, rather than elaborating how an individual functions within the family context (e.g., parent-child or marital relations), biologists are concerned with how one species grows and declines within the context of other species and the environment (e.g., foxes and rabbits). Ecologists, for instance, aim to understand the distribution and abundance of animals and plants, with particular focus on how a species' short- and long-term population dynamics are affected through interaction with the surrounding environment and presence and/or absence of other organisms (e.g., predators, prey). Since early in the 20th century, ecologists have been using a combination of theory, observational data, and mathematical modeling to understand how both intra-specific (within a species) and inter-specific (between-species) interactions contribute to population change.

Two mathematical biologists, Lotka (1925) and Volterra (1926) introduced a robust framework for translating theories about interspecies associations into a mathematical form. This framework has been very useful for modeling, exploring, and theorizing about the distribution and abundance of multiple species living in the same space—symbiosis (see Paracer and Ahmadjian 2000). Formally, a two-species model can be written using a set of nonlinear differential equations. Using x to represent a characteristic of one species (e.g., the size of the population), and y to represent a characteristic of the other species (e.g., also, the size of the population), the dynamics (i.e., the rates of change) of these characteristics can be modeled as

$$\begin{aligned} \frac{dx}{dt} &= a_1 + b_1 x \pm f_1(y, x) \\ \frac{dy}{dt} &= a_2 + b_2 y \pm f_2(x, y) \end{aligned} \tag{2.1}$$

where $\frac{dx}{dt}$ and $\frac{dy}{dt}$ are the *rates of change* in x and y, respectively, over time. The parameters a_1 and a_2 represent each species' "uninfluenced" trajectory when isolated from self-influence and influence of other aspects of the environment, and can be interpreted as an indicator of the species' *homeostatic* or *dynamic equilibrium*. Non-zero values indicate the natural trajectory of change (moving vs. resting). The parameters b_1 and b_2 indicate the extent to which each species (x and y) changes as a function of their own characteristics (e.g., the size of the population at a given moment). Positive parameters indicate a *self-exciting* mechanism wherein higher population levels foster more growth. In contrast, negative b_1 and b_2 parameters indicate a *self-inhibiting* mechanism wherein higher population levels invoke declines. For example, animal populations (humans excluded) tend to keep themselves in check when living in a resource limited environment. The functions f_1 and f_2 articulate how each party changes as a function of the interactions between the two species, and are interpreted as indicators of the *co-regulatory* mechanisms operating within the system (e.g., cooperation/competition). One set of straightforward influence functions can be operationalized as the interactions between the two species (literally the product),

$$\begin{aligned} f_1(y, x) &= \pm\, c_1 yx, \quad \text{and} \\ f_2(x, y) &= \pm\, c_2 xy, \end{aligned} \tag{2.2}$$

where the signs of the parameters c_1 and c_2 indicate how the cross-species interactions affect the growth or decline of each species. For example, when foxes (x) have the opportunity to meet with rabbits (y), species x tends to eat species y, and the interactions invoke positive progression for the foxes ($c_1 > 0$), but negative progression for rabbits ($c_2 < 0$). The influence functions characterize the *predator-prey* nature of the relationship. The species co-regulate, in that the abundance or scarcity of rabbits influences foxes ability to reproduce, and the abundance or scarcity of foxes influences the ability of rabbits to reproduce. Co-regulation may also manifest in ways that are positive for both species. For example, when humming-birds (x) have the opportunity to meet with flowers (y), species x pollinates species y and species y feeds species x. The exchange invokes positive progression for the humming-birds ($c_1 > 0$) and the flowers ($c_2 > 0$). Here, the influence functions characterize the *mutualistic* nature of the co-regulation. Generally, the overall pattern of parameters (e.g., +/+, +/0, +/−, −/−) provides a system-level description of the mutualistic, commensal, parasitic, or competitive nature of the association. Using variants of the Lotka-Volterra equations, biologists have spawned a vast knowledge of how ecological systems change over the short-term and evolve over the long-term (Murray 2002). From a modeling perspective, the elegance of the framework holds great value for articulating and describing aspects of the human ecology.

Applications to Family (and Work-Family) Systems

In combing through the existing literature on family systems we found relatively few papers that explicitly used the biological systems terminology to describe family (or work-family) systems. Noting that the conceptual underpinnings of family systems theory were developed from frameworks used to describe natural systems (e.g., Bowen 1978), Noone (1988) elaborates in detail how the concepts of symbiosis are used in the mental health field and the life sciences, and how they might apply to human families. Werbel and Walter (2002) use the biological analogies of mutualism and parasitism to elaborate on work and family role dynamics. They note how, in a dysfunctional system, an individual's work role may have a negative impact on his or her family role (+/− parasitism), while in a more functional system, feelings of productivity at work may contribute to a better family life and, vice versa, a better family life may contribute to greater productivity at work (+/+ mutualism). Less flattering usage of the biological vernacular is the labeling of young people (up to about age 30) who, after finishing formal schooling and obtaining employment, continue living with their parents as "parasite singles" (Yieh et al. 2004).

Although the biological terminology surrounding symbiosis appears to have been used in a very limited manner in studies of family systems, the accompanying mathematical framework (Eqs. 1 and 2) has proved quite useful in the study of marital relations. Gottman et al. have a long line of work using the theoretico-mathematical framework outlined above to study marital behavior (e.g., Gottman et al. 1999; 2002a, 2002b). Drawing directly from mathematical biology (e.g., Murray 2002), they formulated a version of the Lotka-Volterra models that matched their theoretical models of how spouses (husband $= x$, wife $= y$) influence each other's behavior within marital interactions. This model was then used to identify specific features within the dynamics of conversation that are indicative of relationship quality and contribute to marriage dissolution. The partners' relative positivity during alternating speaking turns within a 15-min conversation/argument (e.g., # of positive comments−# of negative comments) was modeled as a combination of "uninfluenced" tendencies (analogous to the system dynamics described by the parameters a_1, a_2 and b_1, b_2 from above) related to the past history of the relationship and personality (or what we called self-exciting and self-inhibiting mechanisms) and "influence" tendencies (co-regulatory mechanisms) related to how the spouses affected one another. At the basic level, the influence functions are straightforward,

$$f_1(y,x) = \pm\, c_1 y, \quad \text{and}$$
$$f_2(y,x) = \pm\, c_2 x,$$

where person x's trajectory is influenced either positively or negatively by the level of positivity person y expressed in her speaking turn, and person y's subsequent trajectory is influenced either positively or negatively by the level of positivity person x expressed in his speaking turn, and so on, back and forth.

A further level of complexity was introduced to accommodate the finding that negative comments are more detrimental than positive comments are supportive. This is accommodated by a bilinear influence function which includes a threshold k, whereby content with positivity below k (e.g., all negative statements) has a specific amount of influence, c_{1a}, while content with positivity above k (e.g., all positive statements) has a different amount of influence, c_{2a}. Formally,

$$f_1(y,x) = c_{1a} y \text{ when } y \le k_1 \text{ and } f_1(y,x) = c_{1b} y \text{ when } y > k_1 \text{ and}$$
$$f_2(x,y) = c_{2a} x \text{ when } x \le k_2 \text{ and } f_2(x,y) = c_{2b} x \text{ when } x > k_2.$$

In practice, both bilinear (2-part spline functions) and ogive (step functions), and combined bilinear-ogive functions are used to articulate a wide variety of theoretically interesting influence functions (see Madhyastha et al. 2011). The mathematical models based on these functions have been used very successfully to obtain qualitative and quantitative descriptions of the dynamics of couples' marital interactions that are both associated with relationship satisfaction and predictive of dissolution (Gottman et al., 2002a).

Two quick notes about how Gottman and colleagues' model is constructed. (1) The model translates a very straightforward theory about how the two members of a dyad influence one another into a mathematical form. In this case, the influence functions operationalize pure *other-focused regulation* and not co-regulation (as we described earlier). That is, dx is only influenced by the other person's behavior (y), rather than by the product of both partners' behavior ($y*x$) as in the biological models above. Given the micro-time scale being examined, this makes sense. (2) The threshold parameter, k, is an important aspect of the system that may differ across individuals and across family members. Each individual has his or her own threshold, and these differences in tolerance for or perception of negativity contribute to how the dynamics of the conversation unfold (Gottman et al. 1999). In sum, the influence function is the key aspect of the model that can be used to operationalize specific theories about how family members influence one another. Expanding the vocabulary of mutual influences beyond the simple forms of commensalism, mutualism, parasitism, and competition noted above (Murray 2002; Stadler and Dixon 2008), many theories can be articulated precisely within the mathematical model.

For example, Felmlee and Greenberg (Felmlee 2007; Felmlee and Greenberg 1999), wrote out a set of influence functions that emphasize the importance the degree of congruence between family members' states plays in system-level functioning. The influence functions were articulated as

$$f_1(y,x) = \pm\, c_1(y-x), \text{ and}$$
$$f_2(x,y) = \pm\, c_2(x-y),$$

with c_1 and c_2 indicating the extent to which differences between partner's states drive each person's behavioral change. In contrast to the Gottman models articulating a pure other-focused regulation based solely on the other person's state (e.g., y), these dyad-focused influence functions explicitly incorporate dyad-level variables (e.g., $y - x$) that quantify the distance between individuals' states. Holding the other parts of the model fixed (e.g., when the as and bs $= 0$), a positive c_1 parameter indicates person x tends to act in ways that decrease the distance between him and his partner's states (e.g., emotions). When the c_2 parameter is also positive both partners' trajectories tend to move toward each other. In contrast, negative parameters indicate divergence of partners over time. And, a parameter of zero indicates a tendency for independent and uninfluenced change.

Note that the specific influence functions used here, which are based on dyad-level differences $(x-y)$ rather than dyad-level products (xy), were conceptualized for specific kinds of behavior, such as the balance between partner's physical (exercise) and social (# of phone calls) activities. Thus, they are not equivalent to the biological models, and likely do not map directly to the taxonomy of interspecies interactions and population dynamics. Instead, Felmlee (2007) purposively used the modeling framework (a, b and c parameters) to describe a rich typology of dyads' potential dynamics; independents, dependents, cooperatives, reactionaries, etc. Fitting these models to data obtained in a micro-longitudinal study of relationship partners' daily emotional states (similar to the data used in our forthcoming example), Ferrer and Steele (2012) confirmed that the hypothesized typology of couples articulated by the model was indeed related to relationship satisfaction and dissolution.

More generally, all the models above translate precise theory about how partner's behavior or congruence between family members' behavior drives the on-going dynamics of the system into a mathematical form that then allows that theory to be tested against empirical data (for alternative frameworks see Boker and Laurenceau 2006; Butner et al. 2005; Butner et al. 2007; Chow et al. 2010; Ferrer and Nesselroade 2003; Sbarra and Ferrer 2006; Steele and Ferrer 2011). Many more influence functions can be formulated and the work easily extended to reflect the many nuances of family interaction (e.g., Liebovitch et al. 2010).

Data: Empirical Example

To illustrate how the nonlinear dynamic modeling framework outlined above can be used to describe both the ongoing dynamics of a family system *and* changes in those dynamics, we collected a set of EMA data from a White, heterosexual, married couple around the time of the birth of their first child. Using nonlinear dynamic models we describe (a) the dynamic processes (e.g., homeostasis, self-focused regulation and dyad-focused regulation) within a dyadic family system, and (b) changes in those dynamics that occurred in conjunction with passage through a developmental transition.

The study of both *family dynamics* and *change in family dynamics* was facilitated by a multiple-time-scale study design (see Nesselroade 1991; Ram and Gerstorf 2009) wherein multiple reports or assessments were obtained over a relatively short span of time (e.g., 3 weeks of EMA) during multiple "bursts" of measurement obtained at more widely spaced intervals (e.g., years of longitudinal panel) or in different stages of development (e.g., before and after the transition to parenthood). Repeated measures obtained within-burst provide for observation of dynamic process (self-stabilization). Comparison across bursts provides for observation of family-level adaptations to the transition (self-organization).

Data Collection. For 23 consecutive days, the two members of the couple reported on their interpersonal interactions as they went about their daily lives. Specifically, immediately after face-to-face interactions (of longer than 5 min) the husband and wife used separate smartphone-based survey applications to provide independent ratings of the interaction (location, duration, quality) and their emotional states (e.g., happy, sad). Then, this family system went through a major developmental transition—and a "natural experiment" was born (a daughter). Ten days after the birth of the child, the couple resumed reporting about their interpersonal interactions for 42 more days. In total the couple provided reports about 103 interactions, 47 during the pre-baby burst (between 1 and 4 reports per day; Median = 2, SD = 0.69), and 56 during the post-baby burst (0 to 2 per day; Median = 2, SD = 0.54). Here, we make use of responses to the items "How HAPPY do you feel?" and "How SAD do you feel?" each rated using a "touch point continuum" slider-type interface that was digitally coded on a 0−100 scale.

Data Preparation. By design, the couple's reports were provided at unequally spaced intervals reflecting the natural occurrence and time spacing of interactions. To facilitate analyses, the data were pre-processed using usual time-series techniques (see Ramsey et al. 2009; Shumway and Stoffer 2006). First, data for the small amount of days on which no reports were provided were interpolated using third-order local splines (1 point per day). Then, the time series of repeated measures from each burst were smoothed using third-order local splines, with knots placed at 1.0 and 1.2 day intervals for the pre-baby and post-baby bursts, respectively, and resampled at 100 equidistant points. The main objective of the pre-processing was to reduce some of the noise in the data while still maintaining a relatively precise picture of the nonlinear progression of each individual's feelings across time. A few general

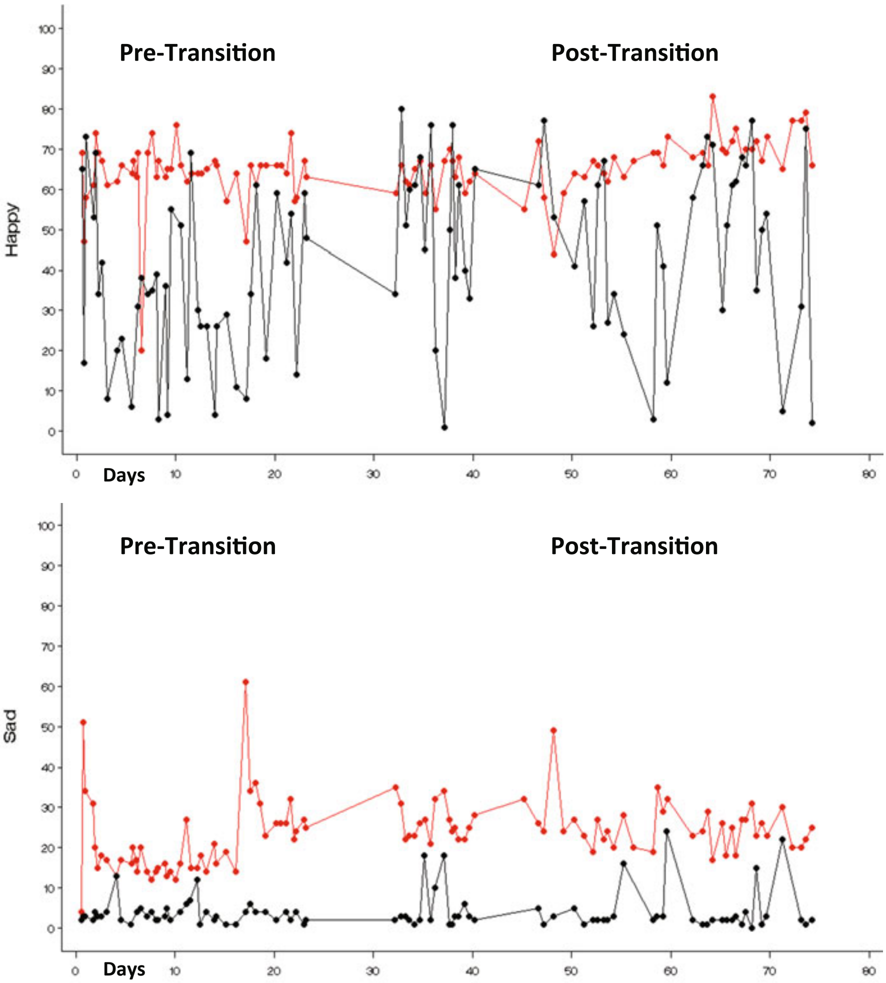


Fig. 2.1 Raw data from husband's (*black*) and wife's (*red*) reports of happy (*top panel*) and sad (*bottom panel*) during two bursts of data collection (pre-baby and post-baby)

features can be noted in Fig. 2.1. Overall, the husband's time series (black lines) was characterized by lower mean level and higher variance in *happy* than the wife's time series (red lines). As well, for both the husband and the wife, there were differences in mean levels and variance between the pre-baby and post-baby periods. While the differences in these dynamic characteristics are interesting in their own right (see Ram and Gerstorf 2009), our main interest here was in describing the dynamics of the family system. Thus, we set these differences aside by standardizing each time-series to a Mean = 0 and SD = 1. This removes the "distraction" of needing to interpret the usual "static" features of the time-series, keeps us focused on the dynamics, and allows for easy comparison of those dynamics across bursts and persons.

Model formulation. Our interest was to use the nonlinear dynamic modeling framework outlined above to describe the dynamics of the dyad's emotions during each burst (pre- and post-baby) and to identify any qualitative differences in those dynamics between the two bursts. Following the Ferrer and Steele (2012) work, we formulated a model based on the Felmlee (2007) influence functions. Specifically, the following model was fit separately to the dyad's *happy* and *sad* time-series for each burst:

$$\frac{dx}{dt} = a_1 + b_1 x \pm c_1(y - x)$$
$$\frac{dy}{dt} = a_2 + b_2 y \pm c_2(x - y)$$

The a_1 (and a_2) parameter indicates the husband's (and wife's) trajectory at equilibrium, when personal and partner's influences are assumed absent. Positive scores on a_1 and a_2 indicate a tendency to remain on an increasing trajectory; negative scores indicate a tendency to remain on a decreasing trajectory; and zero scores indicate a tendency to remain at rest. The b_1 (and b_2) parameter indicates the extent to which the husband's (and wife's) trajectory is prone to change from endogenous influences. Positive scores indicate a tendency for escalation or movement away from equilibrium; negative scores indicate a tendency for inhibition or movement towards the equilibrium. The c_1 (and c_2) parameter indicates the extent to which husband's (wife's) trajectory is influenced exogenously by the imbalance between partners' emotional states (e.g., levels of happy). A positive parameter indicates a tendency to regulate or be regulated toward the partner's score, and a negative parameter indicates a tendency to push or be pushed further away from the partner's score. Note that because the influence function is based on the difference score between the two people, it is not possible to infer whether one or the other member of the dyad is doing the "pulling" or "pushing". Rather, these parameters should be interpreted as family-level parameters that describe the dyadic system as a whole rather than a sum of the parts (Minuchin 1985).

Model fitting. Four models [2 bursts (pre-baby, post-baby) x 2 emotions (happy, sad)] were fitted to time-series data using SAS PROC MODEL, a flexible and general purpose procedure for fitting systems of non-linear differential equations to time-series data and/or simulating data from those equations (e.g., Erdman and Morelock 1996). In brief, the procedure numerically integrates the system of equations using a variable order, variable step-size, backwards difference scheme, and then fits the integral to the data using full information maximum likelihood (or another chosen estimation method) to account for dependence in the error structure due to the simultaneous equation modeling (see SAS documentation and Steele et al. (in press), for further details). Of note, given the didactic nature of our example we have not evaluated the statistical tests of whether each parameter is significantly different than zero, made comparisons among competing models (using model fit indices), nor evaluated the specific influences our arbitrary choice of 100-occasion (vs. 50-occasion) resampling within burst has on the results. Rather, we use the example to demonstrate how the general framework of nonlinear dynamic models might be used to articulate family systems theory.

Results and Interpretation. Estimated parameters from the four models are reported in Table 2.1. Here we provide a detailed interpretation of the dynamics of the dyad's happiness during the pre- and post-baby periods (top section of Table 2.1). The parameters describing the husband's dynamics were $a_1 = -0.006$, $b_1 = -1.474$, and $c_1 = -0.971$. The near zero a_1 parameter indicates that the husband tended toward a non-changing, near zero, resting-state equilibrium. The large negative b_1 parameter indicates a strong self-focused regulatory tendency, such that deviations above or below equilibrium induced a change towards his set-point. The negative c_1 parameter indicates the tendency for dyad-focused regulation; specifically, a tendency to move away from his wife's mood. Following occasions when the wife is happier, $(y - x) > 0$, the husband's happiness would decrease. In complement, on the occasions when the wife is less happy, $(y - x) < 0$, the husband's happiness would tend to increase, perhaps, in the attempt to cheer her up. In comparison, parameters describing the wife's dynamics were estimated to be $a_2 = -0.011$, $b_2 = -0.539$, and $c_2 = +0.509$. As before, the near zero a_2 parameter indicated that the wife tended toward a non-changing, near zero, resting state equilibrium. The negative b_2 parameter indicates a self-regulatory tendency, such that deviations above or below equilibrium induced change towards her set-point. The positive c_2 parameter indicates a tendency to move toward the husband's mood. If the husband was happier than the wife, $(x - y) > 0$, the wife's happiness would tend to increase *and* vice versa. If the husband was less happy than the wife, $(x - y) < 0$, the wife's happiness would tend to decrease.

After the developmental transition, the dynamics were different. The parameter describing the husband's equilibrium state stayed near zero, $a_1 = -0.006$. The self-regulatory parameter substantially decreased, $b_1 = -0.260$, indicating a weaker self-focused regulatory mechanism following childbirth. The influence parameter changed sign to positive and decreased in magnitude, $c_1 = +0.232$, indicating a switch in the husband's dyad-focused regulation–a tendency to move toward (rather than away from) his wife's mood. The parameters describing the wife's dynamics are also somewhat different. The near-zero equilibrium is similar, $a_2 = +0.011$, but there is a stronger self-focused regulatory tendency, $b_2 = -0.905$, and, perhaps in coordination with the husband, a complementary change in sign for $c_2 = -0.500$. In the post-baby period, the wife has the tendency to move away from the husband's mood, perhaps either to retain happiness of equilibrium or try to cheer him up.

The estimated model parameters governing the dynamics of happiness during the pre- and post-baby bursts were used to simulate scenarios of dyadic reactions under a variety of hypothetical perturbations (e.g., emotions perturbed far away from equilibrium, close to equilibrium). One of these many scenarios is visually depicted in the upper panel of Fig. 2.2. In the left, pre-baby panel, we see the dyad progressing along at equilibrium until time = 0, when an event (external or internal) perturbs the system driving the partners into different states (husband lower than equilibrium, wife higher than equilibrium). After the perturbation we see how the coordinated pulling and pushing of self-focused and dyad-focused influences bring the dyad back to homeostasis in about 4 time steps (arbitrary scaling). As can be seen in the right panel, the same perturbation in the post-baby period emphasizes the change in the dynamics. In the post-baby period, the dyad does not reach mutual equilibrium until about the seventh time step.

Table 2.1 Parameter estimates from nonlinear dynamic model articulating (*a*) homeostasis, (*b*) self-focused regulation, and (*c*) dyad-focused regulation

	Husband	Estimate	SE	Wife	Estimate	SE	Husband	Estimate	SE	Wife	Estimate	SE
	Pre-baby: SAD						Post-baby: SAD					
	a_1	−0.006	0.333	a_2	−0.011	0.309	a_1	−0.022	0.196	a_2	0.041	0.124
	b_1	−1.474	0.312	b_2	−0.539	0.401	b_1	−0.260	0.380	b_2	−0.905	0.223
	c_1	−0.971	0.261	c_2	0.509	0.512	c_1	0.232	0.269	c_2	−0.500	0.163
R^2	0.73			0.66			0.61			0.80		
Log-likelihood	−152.9						−156.1					
	Pre-baby: SAD						Post-baby: SAD					
	a_1	0.379	0.374	a_2	0.282	0.341	a_1	−0.007	0.183	a_2	−0.029	0.175
	b_1	−0.497	0.541	b_2	−1.04	0.56	b_1	0.091	0.187	b_2	−1.077	0.32
	c_1	−0.039	0.411	c_2	−0.224	0.637	c_1	0.33	0.141	c_2	−0.505	0.259
R^2	0.81			0.71			0.79			0.69		
Log-likelihood	−135.2						−145.8					

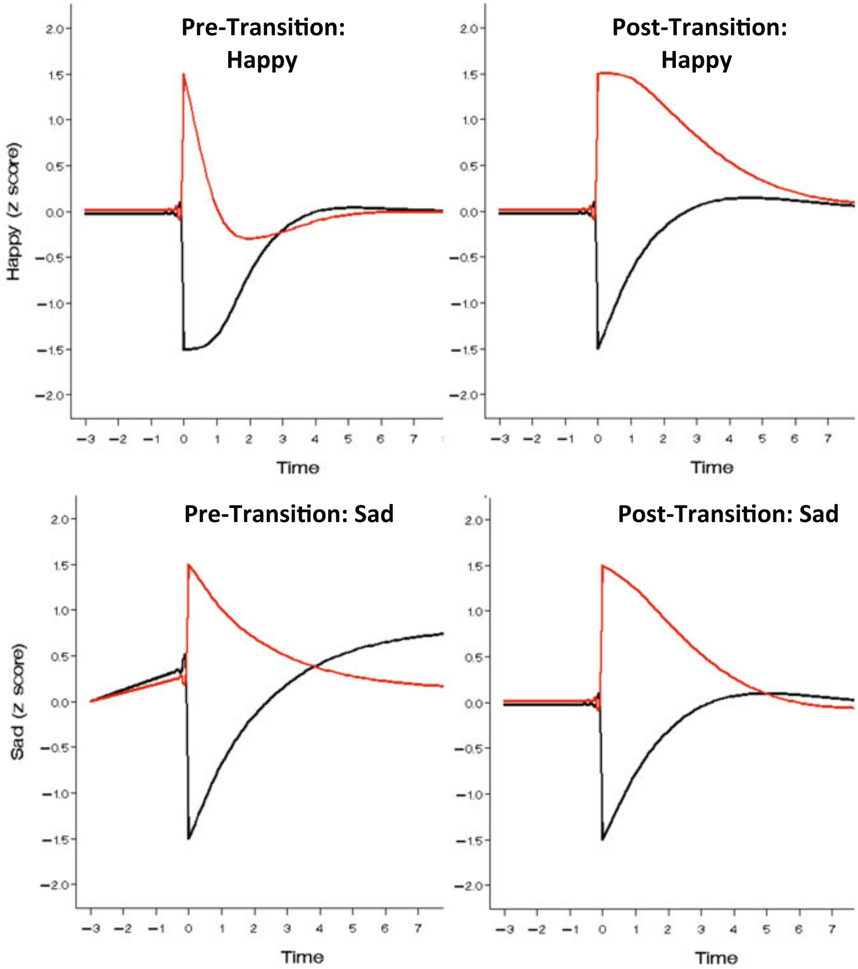


Fig. 2.2 Response simulations of husband's (*black*) and wife's (*red*) changes in happy (*top panel*) and sad (*bottom panel*) during the pre-baby and post-baby periods

Estimated parameters from the pre- and post-baby models of the dynamics of sadness are reported in the lower section of Table 2.1. Briefly, the pre-baby period was characterized by a tendency for increases in sadness for both the husband ($a_1 = 0.379$) and the wife ($a_2 = 0.282$) and seen as upward sloping lines prior to the perturbation in the lower panel of Fig. 2.2. However, there was also a good amount of self-regulatory pressure in the system. When sadness was higher, both partners down-regulated towards the personal equilibrium ($b_1 = -0.479$ and $b_2 = -1.04$). Influence of the impact of differences in levels of sadness between partners on the rate of change was evident for the wife ($c_2 = -0.224$) but not for the husband

($c_1 = -0.039$). The dynamics were somewhat different in the post-baby period. The dyad's homeostasis no longer included any evidence of upward trajectories of sadness. The uninfluenced trajectories had shifted towards zero for both partners ($a_1 = -0.007$ and $a_2 = -0.029$). As well, the husband's self-regulation was weaker ($b_1 = 0.091$) than in the pre-baby period, whereas the wife's self-regulation remained strong and negative ($b_2 = -1.077$), indicating a tendency to return to her homeostasis. Finally, the effect of mood differences on the husband's tendency to change became positive ($c_1 = 0.33$), while the wife's continued to be negative and doubled in magnitude ($b_1 = -0.505$). We briefly note that individuals' emotion regulation is often described as a process that depends on individual set points and homeostatic strategies (roughly corresponding to a_1, a_2) (Forgas and Ciarrochi 2002) regulatory skill and/or endogenous influences (b_1, b_2) (Larsen and Prizmic 2005), and sensitivity to environmental cues (such as empathy based on others' emotions) and/or exogenous influences (c_1, c_2) (Gross and Thompson 2005). Thus, the mathematical models may provide a robust framework for integrating knowledge about the specific phenomena with family systems theory.

Discussion

Systems theory is a useful epistemological approach to study small and large systems, such as family units, that naturally arise in biological and social environments. In biology, systems theory has been used successfully as both a theoretical foundation and to formulate methods for modeling a wide variety of individual-, population-, and community-level phenomena (Begon et al. 1996; Murray 2002). In social sciences, many theorists have noted the parallels between biological units and social units and elaborated how individuals function within context and have direct and indirect impacts on each other's actions and feelings (e.g., Ford and Lerner 1992; Gottlieb 1997; Sameroff 1994; Schneirla 1957; Werner 1957). As technological advances in mobile computing provide for easy collection of intensive longitudinal data (e.g., via smartphones) and advances in computing speed provide capacity for numerical estimation of highly complex models, there are many new opportunities to study family dynamics. The goal of this chapter was to review the basic framework of nonlinear dynamic models that is being used in the ecology (following Wohlwill's 1991, recommendation) and illustrate how it can be used to articulate and model intra-family dynamics.

Theory, Method, and Data

For many decades, systems theory was far ahead of the available methodology. We simply did not have the data or statistical procedures needed to render the theory operational. In recent decades, both barriers have been successfully overcome. Technological innovations allow for frequent (and often unobtrusive) assessment of humans behavior, physiology, and psychological states in natural settings (Bolger et al. 2003; Stone et al. 2007). As illustrated here, the many-occasion, multivariate time-series data suitable for studying the dynamics of change can be obtained. Many of the analytical tools needed to estimate nonlinear dynamic models are now available, even within typically used statistical software (e.g., PROC MODEL in SAS, CollocInfer package in R) (Ramsay et al. 2007).

In fact, these tools, and the rates at which social network and other on-line data are obtained suggest that the data collection methods and analytical methods are now ahead of theory. In the last decade, a few groups have demonstrated the viability of using the precision of the mathematical framework underlying nonlinear dynamic systems for both theorizing and studying family systems (e.g., work by Boker, Gottman, Felmlee, Ferrer, Laurenceau). The results from these efforts are exciting, especially in that their utility for making long-term predictions about tangible and important outcomes (e.g. family dissolution) has been convincingly demonstrated.

Research Questions

Our exploration of theory and empirical modeling highlight how integration of dynamic systems principles and dynamic modeling leads into new directions for basic and applied family research. As seen in our empirical example, the framework reveals a detailed landscape of family change and captures several levels of transition. Using data collected in a multiple-time-scale study design, we were able to model both family dynamics during a specific micro-time frame (within the pre-baby period) and change in family dynamics across a developmental transition (pre- vs. post-baby). More generally, the dynamic systems framework is extremely useful in that it allows for modeling change over many time scales, from change across the course of a single conversation (á la Gottman) to change across the course of a relationship or the entire life span of a family system. As well, the framework can be used to model many different types of family systems and behaviors. We used two coupled equations to model the dynamics of specific emotions within a dyad (á la Felmlee and Ferrer), but the model is easily extended to include more individuals (by adding more equations) or adapted for use with other behaviors (by incorporating different influence functions). That the framework is used by ecologists to model system-level changes in animal populations suggests that it may also have utility for demographers modeling demographic changes in family structure. In sum, we see that the dynamic systems framework reviewed here might be used to investigate a wide variety of family-oriented research questions.

Research Methods

Our exploration of theory and empirical modeling also highlighted some of the ways that the basic modeling framework might be extended. For example, an interesting and important extension would be to use the model to predict relational transitions.

Repeated sampling bursts over longer time periods may reveal shifts in the dynamics (attractor patterns) that may have important implications for tangible outcomes, including relationship formation and dissolution, risk of maltreatment and, more generally family health and well-being. Identifying shifts may be particularly informative in marital or family therapy in terms of identifying the specific events that are strong triggers for positive or negative changes in a couple or in family dynamics. From a modeling perspective, this requires modeling how the parameters describing the dynamics (attractors) change over time. In our empirical example, we modeled the pre- and post-baby periods separately to examine a qualitative change in the parameters. Other forms of dynamic modeling (e.g., state-space grids) have also been used to model qualitative shifts in individual and dyadic development over time (Granic et al. 2003; Lewis et al. 2004; van Dijk and van Geert 2007). With some work, all of these methods might be extended to include time-varying parameters, so as to continuously track how the parameters describing the family dynamic evolve over time, and or shift in relation to specific events (Chow et al. 2011; Molenaar and Ram 2010; Tan et al.2011). Such extensions might then be used prospectively to identify periods of transition and/or risk in romantic relationships, parental functioning, friendships, and other systems. Not yet known, though, are the number and spacing of assessments needed for such evaluations (for discussion, see Collins 2006; Ram and Gerstorf 2009; Shiyko and Ram 2011).

Our empirical example highlighted how nonlinear dynamic models can be applied to time-series data to describe the dynamics of a dyadic system. Data from multiple systems would offer additional opportunities for describing inter-system (e.g., between-couple) similarities and differences that might also be examined in relation to other inter-system differences (e.g., demographics, relationship history). Placing the models within a multilevel modeling framework that allows for simultaneous examination of within-system change and between-system differences is relatively straightforward (Boker and Laurenceau 2006; Chow et al. 2010). Given that a primary area of interest for family researchers and practitioners is understanding why maladaptive relationships are resistant to change, such an approach may shine additional light on patterns of resistance and potential for intervention.

Conclusion

The success of the modeling approach is rooted in the explicit articulation of families as dynamic entities that interact in complex ways—as systems. Nonlinear dynamic models, such as those used here, focus on *change*, rather than focusing on where a person or family happens to be located at a particular or arbitrary moment in time. The "outcomes" in the models are derivatives, rates of change and acceleration, and thus focus inquiry on the dynamics of change. Individuals in the same family are modeled simultaneously, recognizing both their organismic continuity and reciprocal relations. These ideas have been discussed in family systems theory for decades. Now we have the tools and opportunity to write them down in a precise mathematical form that allows them to be tested against empirical data. We look forward to what emerges!

Acknowledgements The authors are grateful for the support provided by the National Institute of Health (Grants RC1-AG035645, R21-AG032379, R21-AG033109, R03-CA171809) and the Penn State Social Science Research Institute.

References

Begon, M., Harper, J. L., & Townsend, C. R. (1996). *Ecology: Individuals, populations and communities, Third edition*. Cambridge: Blackwell Science.

Boker, S. M., & Laurenceau, J. P. (2006). Dynamical systems modeling: An application to the regulation of intimacy and disclosure in marriage. In T. A. Wall, & J. L. Schafer (Eds.), *Models for intensive longitudinal data* (pp. 195–218). New York: Oxford University Press.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54,* 579–616. doi:10.1146/annurev.psych.54.101601.145030.

Bowen, M. (1978). *Family therapy in clinical practice*. New York: Jason Aronson.

Butner, J., Amazeen, P. G., & Mulvey, G. M. (2005). Multilevel modeling of two cyclical processes: Extending differential structural equation modeling to nonlinear coupled systems. *Psychological Methods, 10,* 159–177. doi:10.1037/1082-989X.10.2.159.

Butner, J., Diamond, L. M., & Hicks, A. M. (2007). Attachment style and two forms of affect coregulation between romantic partners. *Personal Relationships, 14*(3), 431–455. doi:10.1111/j.1475-6811.2007.00164.x.

Chow, S.-M., Haltigan, J. D., & Messinger, D. S. (2010). Dynamic infant-parent affect coupling during face-to-face/still-face. *Emotion, 10,* 101–114.

Chow, S.-M., Zu, J., Shifren, K., & Zhang, G. (2011). Dynamic factor analysis models with time-varying parameters. *Multivariate Behavioral Re*search, *46*(2), 303–339.

Collins, L. M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model. *Annual Review of Psychology, 57,* 505–528.

Cox, M. J., & Paley, B. (1997). Families as systems. *Annual Review of Psychology, 48,* 243–267.

Erdman, D., & Morelock, M. M. (1996). A study of kinetics: The estimation and simulation of systems of first-order differential equations. Proceedings of the 21st Annual SAS Users Group International Conference, 1407–1414. http://support.sas.com/rnd/app/papers/kinetics.pdf.

Felmlee, D. H. (2007). Application of dynamic systems analysis to dyadic interactions. In A. D. Ong & M. van Dulmen (Eds.), *Handbook of methods in positive psychology* (pp. 409–422). New York: Oxford University Press.

Felmlee, D. H., & Greenberg, D. F. (1999). A dynamic systems model of dyadic interaction. *Journal of Mathematical Sociology, 23,* 155–180.

Ferrer, E., & Nesselroade, J. R. (2003). Modeling affective processes in dyadic relations using dynamic factor analysis. *Emotion, 3,* 344–360.

Ferrer, E., & Steele, J. (2012). Dynamic systems analysis of affective processes in dyadic interactions using differential equations. In G. R. Hancock & J. R. Harring (Eds.), *Advances in longitudinal methods in the social and behavioral sciences* (pp. 111–134). Charlotte: Information Age Publishing.

Fogel, A. (1993). Two principles of communication: Co-regulation and framing. In J. Nadel & L. Camaioni (Eds.), *New perspectives in early communicative development* (pp. 9–22). London: Routledge.

Ford, D., & Lerner, R. (1992). *Developmental systems theory: An integrative approach*. Newbury Park: Sage.

Forgas, J. P., & Ciarrochi, J. V. (2002). On managing moods: Evidence for the role of homeostatic cognitive strategies in affect regulation. *Personality and Social Psychology Bulletin, 28*(3), 336–345.

Gottlieb, G. (1997). Commentary: A systems view of psychobiological development. In D. Magnusson (Ed.), *The lifespan development of individuals: Behavioral, neurobiological, and psychosocial perspectives: A synthesis* (pp. 76–103). New York: Cambridge University Press.
Gottman, J., Swanson, C., & Murray, J. (1999). The mathematics of marital conflict: Dynamic mathematical nonlinear modeling of newlywed marital interaction. *Journal of Family Psychology, 13,* 3–19.
Gottman, J. M., Murray, J. D., Swanson, C. C., Tyson, R., & Swanson, K. R. (2002a). *The mathematics of marriage: Dynamic nonlinear models*. Cambridge: MIT Press.
Gottman, J., Swanson, C., & Swanson, K. (2002b). A general systems theory of marriage: Nonlinear difference equation modeling of marital interaction. *Personality and Social Psychology Review, 6,* 326–340.
Granic, I. (2005). Timing is everything: Developmental psychopathology from a dynamic systems perspective. *Developmental Review, 25,* 386–407. doi:10.1016/j.dr.2005.10.005.
Granic, I., Hollenstein, T., Dishion, T. J., & Patterson, G. R. (2003). Longitudinal analysis of flexibility and reorganization in early adolescence: A dynamic systems study of family interactions. *Developmental Psychology, 39,* 606–617.
Gross, J. J., & Thompson, R. A. (2005). Emotion regulation: Conceptual foundations. In J. J. Gross (Ed.), *Handbook of emotion regulation* (pp. 3–24). New York: Guilford Press.
Larsen, R. J., & Prizmic, Z. (2005). Affect regulation. In L. F. Barrett, P. M. Niedenthal, & P. Winkielman (Eds.), *Emotion and consciousness* (pp. 40–61). New York: Guilford Press.
Lewis, M. D., Zimmerman, S., Hollenstein, T., & Lamey, A. V. (2004). Reorganization in coping behavior at 1½ years: Dynamic systems and normative change. *Developmental Science, 7,* 56–73.
Liebovitch, L. S., Vallacher, R. R., & Michaels, J. (2010). Dynamics of cooperation-competition interaction models. *Peace and Conflict, 16,* 175–188.
Lotka, A. J. (1925). *Elements of physical biology*. Baltimore: Williams & Wilkins Co.
Lunkenheimer, E. S., & Dishion, T. J. (2009). Developmental psychopathology: Maladaptive and adaptive attractors in children's close relationships. In S. Guastello, M. Koopmans, & D. Pincus (Eds.), *Chaos and complexity: Recent advances and future directions in the theory of nonlinear dynamical systems psychology* (pp. 282–306). New York: Cambridge University Press.
Lunkenheimer, E. S., Hollenstein, T., Wang, J., & Shields, A. M. (2012). Flexibility and attractors in context: Family emotion socialization patterns and children's emotion regulation in late childhood. *Nonlinear Dynamics, Psychology, and Life Sciences, 16,* 269–291.
Madhyastha, T. M., Hamaker, E. L., & Gottman, J. M. (2011). Investigating spousal influence using moment-to-moment affect data from marital conflict. *Journal of Family Psychology, 25,* 292–300.
Minuchin, S. (1974). *Families and family therapy*. Cambridge: Harvard University Press.
Minuchin, P. (1985). Families and individual development: Provocations from the field of family therapy. *Child Development, 56,* 289–302.
Molenaar, P. C. M., & Ram, N. (2010). Dynamic modeling and optimal control of intra-individual variation: A computational paradigm for non-ergodic psychological processes. In S. M. Chow, E. Ferrer, & F. Hsieh (Eds.), *Statistical methods for modeling human dynamics: An interdisciplinary dialogue* (pp. 13–35). New York: Routledge.
Murray, J. D. (2002). *Mathematical biology: I. An introduction,* (3rd ed.). New York: Springer.
Nesselroade, J. R. (1991). The warp and woof of the developmental fabric. In R. Downs, L. Liben, & D. Palermo (Eds.), *Visions of development, the environment, and aesthetics: The legacy of Joachim F. Wohlwill* (pp. 213–240). Hillsdale: Erlbaum.
Noone, R. J. (1988). Symbiosis, the family, and natural systems. *Family Process, 27,* 285–292.
Paracer, S., & Ahmadjian, V. (2000). *Symbiosis: An introduction to biological association* (2nd ed.). New York: Oxford University Press.
Ram, N., & Gerstorf, D. (2009). Time structured and net intraindividual variability: Tools for examining the development of dynamic characteristics and processes. *Psychology and Aging, 24,* 778–791.

Ram, N., & Nesselroade, J. R. (2007). Modeling intraindividual and intracontextual change: Operationalizing developmental contextualism. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 325–342). Mahwah: Lawrence Erlbaum.
Ram, N., & Pedersen, A. B. (2008). Dyadic models emerging from the longitudinal structural equation modeling tradition: Parallels with ecological models of interspecifc interactions. In N. A. Card, J. P. Selig, & T. D. Little (Eds.), *Modeling dyadic and interdependent data in the developmental and behavioral sciences* (pp. 87–105). New York: Routledge.
Ramsay, J. O., Hooker, G., Cao, J., & Campbell, D. (2007). Parameter estimation in ordinary differential equations: A generalized smoothing approach. *Journal of the Royal Statistical Society, 69,* 741–796.
Ramsey, J. O., Hooker, G., & Graves, S. (2009). *Functional data analysis with R and Matlab*. New York: Springer.
Sameroff, A. (1983). Developmental systems: Contexts and evolution. In W. Kessen (Ed.), *Handbook of child psychology Vol. 1: History, theory and methods* (4th ed., pp. 237–294). New York: Wiley.
Sameroff, A. (1994). Developmental systems and family functioning. In R. D. Parke & S. G. Kellman (Eds.), *Exploring family relationships with other social contexts*. (pp. 199–214). Hillsdale: Erlbaum.
Sbarra, D. A., & Ferrer, E. (2006). The structure and process of emotional experience following nonmarital relationship dissolution: Dynamic factor analyses of love, anger, and sadness. *Emotion, 6,* 224–238.
Schneirla, T. C. (1957). The concept of development in comparative psychology. In D. B. Harris (Ed.), *The concept of development: An issue in the study of human behavior* (pp. 78–108). Minneapolis: University of Minnesota Press.
Shiyko, M. P., & Ram, N. (2011). Conceptualizing and estimating process speed in studies employing ecological momentary assessment designs: A multilevel variance decomposition approach. *Multivariate Behavioral Research, 46,* 875–899.
Shumway, R. H., & Stoffer, D. S. (2006). *Time series analysis and its applications: With R examples* (2nd ed.). New York: Springer.
Stadler, B., & Dixon, T. (2008). *Mutualism: Ants and their insect partners*. New York: Cambridge University Press.
Steele, J. S., & Ferrer, E. (2011). Latent differential equation modeling of self-regulatory and coregulatory affective process. *Multivariate Behavioral Research, 46,* 956–984.
Steele, J. S., Ferrer, E. & Nesselroade, J. R. (in press) An idiographic approach to estimating models of dyadic interactions with differential equations. *Psychometrika*.
Stone, A., Shiffman, S., Atienza, A. A., & Nebeling, L. (2007). *The Science of real-time data capture: Self-reports in health research*. New York: Oxford University Press.
Tan, X., Shiyko, M. P., Li, R., Li, Y., & Dierker, L. (2011). Intensive longitudinal data and the time-varying effects model. *Psychological Methods, 17*(1), 61–77.
Thelen, E., & Smith, L. B. (1998). Dynamic systems theories. In W. Damon (Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (5th ed., pp. 563–633). New York: Wiley.
Thelen, E., & Ulrich, B. D. (1991). Hidden skills: A dynamic systems analysis of treadmill stepping during the first year. *Monographs of the Society for Research in Child Development, 56*(1), 1–103.
van Dijk, M., & van Geert, P. (2007). Wobbles, humps and sudden jumps: A case study of continuity, discontinuity and variability in early language development. *Infant and Child Development, 16,* 7–33.
van Geert, P. (1998). A dynamic systems model of basic developmental mechanisms: Piaget, Vygotsky, and beyond. *Psychological Review, 105*(4), 634–677. doi:10.1037/0033-295X.105.4.634-677.
von Bertalanffy, L. V. (1968). *General systems theory*. New York: George Braziller.

Volterra, V. (1926). Fluctuations in the abundance of a species considered mathematically. *Nature, 118,* 558–560.

Werbel, J., & Walter, M. H. (2002). Changing views of work and family roles: A symbiotic perspective. *Human Resource Management Review, 12,* 293–298.

Werner, H. (1957). The concept of development from a comparative and organismic view. In D. B. Harris (Ed.), *The concept of development: An issue in the study of human behavior* (pp. 125–148). Minneapolis: University of Minnesota Press.

Wohlwill, J. F. (1991). Relations between method and theory in developmental research: A partial-isomophism view. In P. Van Geert & L. P. Mos (Eds.), *Annals of theoretical psychology* (Vol. 7, pp. 91–138). New York: Plenum Press.

Yieh, K., Tsai, Y.-F., & Kuo, C.-Y. (2004). An exploratory study of "parasites" in Taiwan. *Journal of Family and Economic Issues, 25,* 527–539.

Chapter 3
Representing Trends and Moment-to-Moment Variability in Dyadic and Family Processes Using State-Space Modeling Techniques

Sy-Miin Chow, Whitney I. Mattson and Daniel S. Messinger

Consider the various forms of change processes that may be encountered in the studies of dyadic and family processes. In some cases, researchers may wish to focus on extracting systematic trends that are "smooth," or unfold in structured ways. Such processes are referred to as *intraindividual trends* in the present chapter and examples may include a child who becomes more distressed as his/her attempts at getting a parent to comply with a request are repeatedly denied, or a couple whose disagreements continue to strain their relationship over the years. In other cases, there may be more subtle changes—termed *intraindividual variability* in the present chapter—that are of interest to the researchers. These may include deviations in affect triggered by efforts aimed at calming the child despite the parents' refusal of the child's request, and temporary alleviation of distress between the couple while on vacation together.

Through the use of different methodological as well as experimental procedures, intraindividual variability has repeatedly emerged as a reliable, quantifiable, and useful source of individual differences. For instance, measures of intraindividual variability have been found to be indicative of between-person differences in cognitive performance (Allaire and Marsiske 2005) emotions (Eid and Diener 1999), family dynamics (Schermerhorn et al. 2010), reactivity to stress and related hormonal changes (Almeida et al. 2009; Sliwinski et al. 2009), as well as susceptibility to neuropsychiatric disorders such as Attention-Deficit/Hyperactivity Disorder (ADHD; Frazier-Wood et al. 2012). Longitudinal data, especially those measured intensively within participants over time (e.g., Bolger et al. 2003; Nesselroade and Ford 1985; Walls and Schafer 2006), provide an initial step toward hastening our understanding

S.-M. Chow (✉)
Department of Human Development and Family Studies, The Pennsylvania State University, University Park, PA, USA
e-mail: symiin@psu.edu

W. I. Mattson · D. S. Messinger
Department of Psychology, University of Miami, Miami, FL, USA
e-mail: wimattson@gmail.com

D. S. Messinger
e-mail: dmessinger@miami.edu

S. M. McHale et al. (eds.), *Emerging Methods in Family Research,*
National Symposium on Family Issues 4, DOI 10.1007/978-3-319-01562-0_3,

of the distinctions between trends and transient intraindividual variability. Unfortunately, when intraindividual variability data are analyzed without regard to the underlying trends, inferences concerning the nature of each family member's dynamics and the interactive influences among family members may be biased. This is further aggravated by the fact that contemporary longitudinal models used to represent change in many areas of social and behavioral sciences are dominated, for the most part, by models aimed at capturing trends, as opposed to intraindividual variability (e.g., Kurdek 2005).

Growth curve and mixed effects modeling approaches (Bryk and Raudenbush 1987; McArdle and Nesselroade 2003) are among the most popular for modeling change in the social and behavioral sciences. While these approaches are commonly used to extract concurrent linkages among growth curve processes, specific notions concerning how one process helps drive the changes in other processes over time are not tested explicitly. In the example of the parent-child dyad noted above, growth curve models can validate that the rise in the child's distress is *related* to the decline in the parent's positive affect over time, and that between-dyad differences in the slopes of these processes can be explained using external covariates such as attachment status. However, growth curve models cannot be used, without some added modifications, to represent how parents and children in different dyads deviate from their own respective growth and decline trajectories over time. The few exceptions that do integrate representation of patterns of intraindividual variability into growth curve models (e.g., Bollen and Curran 2004; Browne 1993) tend to emphasize the trend-based component within a single growth process, as opposed to the intraindividual variability structures observed across multiple processes. While standard mixed effects and growth curve models provide the option to use specific error structures to define the properties of the residuals (e.g., time dependency in the form of autoregressions among residuals; Verbeke and Molenberghs 2000), most commercial software for mixed effects models do not allow enough flexibility for users to specify intraindividual variability within trend-based models. In addition, most growth curve models assume that individuals' change trajectories are deterministic, namely, knowing the true levels at a previous time point provides complete information concerning future levels; no further within-person deviations from the designated change trajectories are allowed other than measurement errors. While this kind of deterministic representation may be appropriate for certain change processes, dyadic and family processes often show more complex intraindividual trends than those posited in standard growth curve models. To this end, trend-based models that allow for additional within-person uncertainties in the form of process noises—termed stochastic models—can serve as a helpful alternative.

Our goal is to elucidate how the state—space modeling framework can be utilized to construct change models targeted toward capturing patterns of intraindividual variability around stochastic trends during a dyadic interaction process. The chapter is organized as follows: We first discuss possible modeling choices for representing trends and demonstrate how these models can be formulated as state–space models. Second, we discuss how other state—space models can be effectively combined with trend-based models to represent changes that ebb and flow around these trends.

The resultant model can thus be used to simultaneously represent more—or—less systematic patterns of growth and decline, as well as the occasional within-person (and relatedly, within-dyad) deviations from the otherwise stable trends.

State–Space Modeling Framework

While many longitudinal studies are based on panel designs, intensive repeated measures designs have become increasingly popular across a variety of disciplines (Bolger et al. 2003; Nesselroade and Ford 1985; Walls and Schafer 2006). A typical data set consists of a large number of repeated measurements (say $T > 50$), while the number of subjects may be as few as, or larger, than one. Handling intensive longitudinal data using contemporary techniques such as structural equation modeling (SEM) can be a cumbersome task (Chow et al. 2010b; Hamaker et al. 2002). In contrast, state–space modeling techniques provide a flexible repertoire of tools for examining intensive longitudinal as well as panel data.

In the linear state–space modeling framework, the general model consists of a dynamic model and a measurement model, expressed respectively as

$$\eta_{it} = \nu + \mathrm{B}\eta_{i,t-1} + \zeta_{it}, \quad \zeta_{it} \overset{\text{i.i.d}}{\sim} \mathrm{N}(0, \Sigma_\zeta) \tag{3.1}$$

$$y_{it} = \tau + \Lambda\eta_{it} + \varepsilon_{it}, \quad \varepsilon_{it} \overset{\text{i.i.d}}{\sim} \mathrm{N}(0,\Sigma_\varepsilon) \tag{3.2}$$

$$\eta_{i1} \sim \mathrm{N}(a, \Sigma_0), \quad t = 1, \ldots, T;\ i = 1, \ldots, n,$$

where Eq. (3.1) is denoted as the dynamic model, which expresses how the latent variables change from time $t-1$ to time t; η_{it}, is a $w \times 1$ vector of latent variables (also called "state variables"), ζ_{it} is a vector of process noise components, ν is a vector of intercepts, and B is a transition matrix that relates the latent variables at time $t-1$ to those at time t. Equation (3.2) is referred to as the measurement model and it dictates the relations among the observed and latent variables. Here, y_{it} is a $p \times 1$ vector of observed indicators, τ is a vector of intercepts, while ε_{it} is a $p \times 1$ vector of unique components (or "measurement errors"). In the state–space context, initial condition (specifically, the mean, a, and covariance matrix, Σ_0) of the latent variables, $\eta_{i,1}$, has to be specified. Such information reflects the distributional properties of the latent variables prior to and up to the first observed time point.

Illustrative Examples

We provide a series of illustrative examples to demonstrate the applicability of state–space modeling techniques to empirical research. In example I, we illustrate the use of selected state–space models to capture slow-varying trends in dyadic data. In example II, we illustrate some possible ways in which other submodels can be added to represent intraindividual variability around those trends.

Data Descriptions

The sample included 20 infants whose older siblings had been diagnosed with Autism Spectrum Disorders (ASD-sibs). ASD-sibs had at least one sibling who was diagnosed with Autism, Asperger's Disorder, or Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS). Also included were 18 infants comparison siblings (COMP-sibs) whose older sibling(s) had not been diagnosed with an ASD and showed no evidence of heightened ASD symptomatology. The illustrative data presented here have been published previously in Chow et al. (2010). The data consisted of mother-infant dyads observed in the Face-to-Face/Still-Face (FFSF) protocol (Adamson and Frick 2003; Tronick et al. 1978), during which mothers were first asked to play with their baby without toys for 3 min (Face-to-Face episode, FF), stop playing and maintain a still face with no emotional expression for 2 min (Still-Face, SF), and resume play for another 3 min (Reunion episode, RE). Ratings of each dyad member's affective valence level were collected using the Continuous Measurement System (CMS; Chow et al. 2010) from 12 student raters, with higher ratings indicating higher (i.e., more positive) valence. After excluding data from dyads who did not complete the FFSF procedure due to persistent distress, a final sample of 36 dyads was used for model fitting purposes.

Prior to model fitting, we discarded the first 10 s of ratings from all raters to minimize the initial delays manifested by some raters as they were getting accustomed to the rating protocol. To minimize idiosyncratic between-rater variations and other arbitrary differences in ratings, we began by standardizing each time series using the means and standard deviations of each rater's ratings during a particular episode. Contrary to the univariate composite scores used in Chow et al. (2010), we constructed three sets of composite scores for each participant by means of item parceling (Kishton and Widaman 1994). In addition, no detrending procedure was adopted prior to model fitting to preserve the systematic trends in the data. A plot of the item parcel scores for infants and mothers from three randomly selected dyads across the three FFSF episodes is shown in Fig. 3.1.

Software for Model Fitting

All model fitting was performed using *SsfPack*, a suite of C routines for implementing state–space modeling techniques, including numerical routines for fitting models in state–space form (Koopman et al. 1999). *SsfPack* is one of the many statistical packages implemented using Ox, an object-oriented matrix programming language (Doornik 1998).

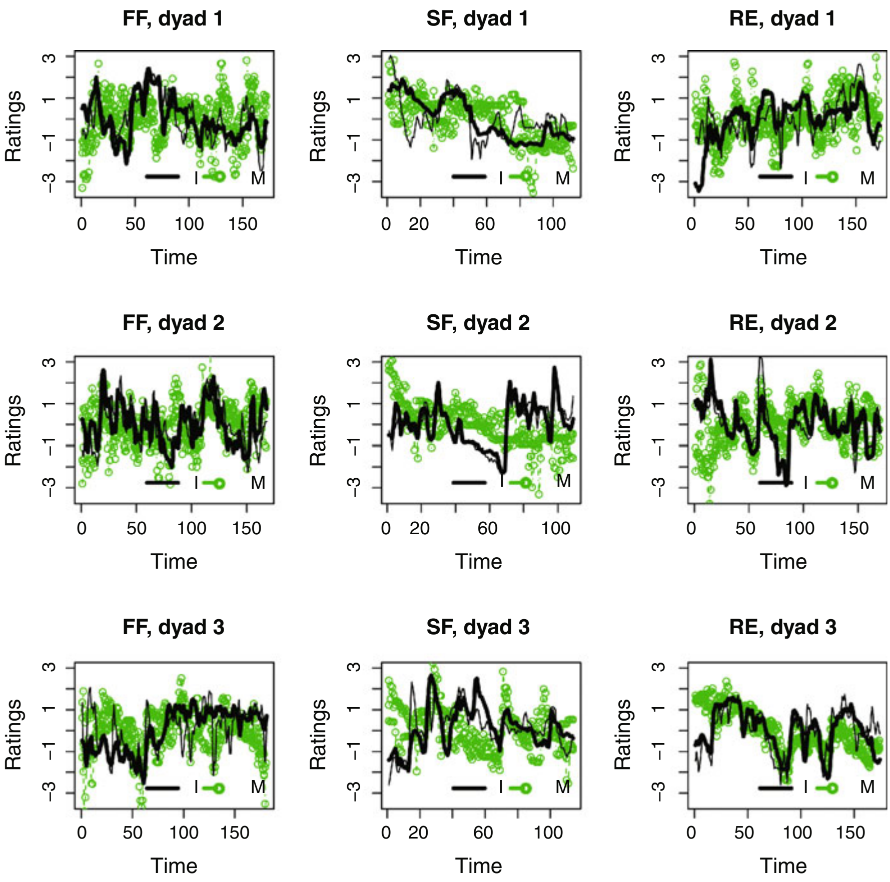


Fig. 3.1 A plot of the item parcel scores for infant (I) and mother (M) from three randomly selected dyads across the three FFSF episodes

Selected Models for Trends

We present several common models for trends in the state–space literature. All of them can be structured as special cases of a model known as the local linear trend (LLT) model (Durbin and Koopman 2001). We first describe properties of the local linear trend model, followed by an illustration of how two other models, a random walk (RW) model and a linear growth curve model, can be obtained as special cases. Due to space constraints, we focus on explicating the theoretical meanings of the components in each model, but only present modeling details and results from the RW model, the model that best describes the characteristics of the present data out of the three models considered.

Local linear trend (LLT) model. The local linear trend model expresses an individual's level on a dependent variable of interest at time t as a function of the local level and the local slope at time $t-1$. Collectively, the local levels and slopes for the two dyad members form a bivariate nonparametric curve written as

$$\begin{bmatrix} \mu_{infant,it} \\ \beta_{infant,it} \\ \mu_{mother,it} \\ \beta_{mother,it} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{infant,i,t-1} \\ \beta_{infant,i,t-1} \\ \mu_{mother,i,t-1} \\ \beta_{mother,i,t-1} \end{bmatrix} + \begin{bmatrix} \zeta_{\mu_{infant},it} \\ \zeta_{\beta_{infant},it} \\ \zeta_{\mu_{mother},it} \\ \zeta_{\beta_{mother},it} \end{bmatrix} \tag{3.3}$$

where $\mu_{infant,it}$ and $\mu_{mother,it}$ represent the local levels of infant and mother in dyad i, respectively, at time t; $\beta_{infant,it}$ and $\beta_{mother,it}$ are the two dyad members' local slopes at time t, while $[\zeta_{\mu_{infant},it}\ \zeta_{\beta_{infant},it}\ \zeta_{\mu_{mother},it}\ \zeta_{\beta_{mother},it}]'$ represents a vector of process noises or within-person deviations in local level and slope that are not accounted for by the previous level and slope at time $t-1$. It is routinely assumed that $\mathrm{Cov}(\zeta_{it}) = \Sigma_\zeta$ across all dyads and time points is a diagonal matrix, namely, no shared covariances are assumed among the process noises of the local level and local slope components. However, this assumption can be relaxed when needed.

A multivariate measurement model can be specified as

$$\begin{bmatrix} Infant_1,it \\ Infant_2,it \\ Infant_3,it \\ Mother_1,it \\ Mother_2,it \\ Mother_3,it \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \lambda_{21} & 0 & 0 & 0 \\ \lambda_{31} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \lambda_{52} & 0 \\ 0 & 0 & \lambda_{62} & 0 \end{bmatrix} \begin{bmatrix} \mu_{infant,it} \\ \beta_{infant,it} \\ \mu_{mother,it} \\ \beta_{mother,it} \end{bmatrix} + \begin{bmatrix} \varepsilon_{infant1,it} \\ \varepsilon_{infant2,it} \\ \varepsilon_{infant3,it} \\ \varepsilon_{mother1,it} \\ \varepsilon_{mother2,it} \\ \varepsilon_{mother3,it} \end{bmatrix} \tag{3.4}$$

where the item parcels for infant and mother in dyad i, $Infant_{1,it}$—$Mother_{3,it}$, are manifest variables included in vector y_{it} to indicate the unobserved local levels, $\mu_{infant,it}$ and $\mu_{mother,it}$, but not the local slopes, $\beta_{infant,it}$ and $\beta_{mother,it}$; λ_{jk} denotes the factor loading of the kth factor on the jth item parcel, with the first parcel loading on each factor fixed at unity for identification purposes.

What do the local levels and local slopes from the LLT model capture? Putting the example of the "nagging" child and the "non-compliant" mother into the context of the SF episode, the local levels indicate the underlying or unobserved affect valence levels of the mother and the infant at each time point. The infant's valence level may decrease over the entire SF episode (e.g., Ekas et al. 2013), but not necessarily at the same rate throughout the entire episode. Alternatively, the infant's valence level may first decrease and then increase or stabilize over time, or in yet other cases, it may fluctuate in a cyclical manner. The local slopes provide a way to approximate all these trajectories because the magnitudes and signs of the slopes can change at each time point due to the process noises or "random shocks" from the terms, $\zeta_{\beta_{infant},it}$ and $\zeta_{\beta_{mother},it}$. In addition, the local levels at each time point can deviate further from the local levels at the previous time point due to the process noises, $\zeta_{\mu_{infant},it}$ and $\zeta_{\mu_{mother},it}$. The process noises or random shocks may be conceived as additional sources of uncertainties or events that can affect the local levels or local slopes at each time point but they themselves do not show continuity over time in a process-oriented manner. For instance, the infant may become distracted by the sight of a new object and temporarily cease crying. This event constitutes a shock that changes the local level and/or local slope of the infant's affective level, but we are not interested in the event itself as a process and thus do not include it as a dependent variable in the model.

The LLT model is, of course, just one of the many possible models for trends. In practice, to fit a state–space model such as the LLT model, a researcher has to first set up a data set in the long format, with T rows of data and p columns of variables for each participant ($p = 6$ in this particular example, corresponding to the six sets of parcel scores for mother and infant as shown in Eq. (3.4)). The next step is to define the state–space model of interest by specifying elements of the vectors, ν, τ and a, and the matrices, B, $\Lambda, \Sigma_{\zeta}, \Sigma_{\varepsilon}$ and Σ_0. For instance, comparing Eq. (3.3) to Eq. (3.1), it can be seen that $\mathrm{B} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ and $\nu =$ a null vector.

Comparing Eq. (3.4) to Eq. (3.2) reveals that $\Lambda = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \lambda_{21} & 0 & 0 & 0 \\ \lambda_{31} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \lambda_{52} & 0 \\ 0 & 0 & \lambda_{62} & 0 \end{bmatrix}$ and τ is a null vector.

Several options exist for specifying the initial condition means in a and the covariance matrix, Σ_0. One option, for instance, is to specify a to be a vector of zeros and Σ_0 to be a diagonal matrix with large variances in its main diagonal. Other options include fixing their elements at known values or estimating them as parameters in the model (e.g., Harvey and Souza 1987; Chow et al. 2010b). Note that unlike the process noise variances for local levels and local slopes, which capture *within-dyad deviations* as observed across all dyads over time, these initial variances may also include pre-existing *between*-dyad differences in initial levels and slopes prior to and up to the first available data point.

Bivariate random walk (RW) model. Another commonly adopted model for trend analysis in the physical sciences and econometric literature, the RW model (Collins and De Luca 1994), can also be obtained as a special case of the LLT model. This can be accomplished by setting all mean and variance parameters for the local slope components to zero. Alternatively, all vector and matrix entries corresponding to the local slopes can simply be omitted from the model (e.g., $\eta_{it} = [\mu_{infant,it}\ \mu_{mother,it}]'$ and B becomes a 2×2 identity matrix). In the context of the FFSF protocol, this means that we are simply allowing the unobserved valence levels of each dyad to deviate from the dyad's previous valence levels at time $t-1$ through direct random shocks imposed by $\zeta_{\mu_{infant},it}$ and $\zeta_{\mu_{mother},it}$ without the random shocks associated with the local slopes. The RW model may be sufficient when the within-partner trends of interest consist primarily of level shifts and less of local peaks and troughs that show ongoing changes in direction.

Linear growth curve model. The well-known linear growth curve model (Meredith and Tisak 1990) can also be structured as a special case of the LLT model. This model posits that the same amount of change is observed in each individual's latent trajectory from one time point to another, thereby giving rise to a linear change trajectory. Interindividual differences are allowed in the intercept (usually defined as the initial level at the first time point) and slope, thus giving rise to a different linear growth trajectory for each individual. However, there are no other sources of within-person deviations from each person's own linear trajectory except for measurement errors.

The linear growth curve model can be formulated as a special case of the local linear trend model by setting Σ_ζ to a null matrix, indicating no process noise at the latent level. Interindividual differences in intercept and slope are incorporated into the initial condition mean and covariance structures as

$$a = [\mu_{intercept,infant} \quad \mu_{slope,infant} \; \mu_{intercept,mother} \quad \mu_{slope,mother}]' \tag{3.5}$$

while Σ_0 contains the associated variances (interindividual differences) in intercepts and slopes for mothers and infants, and pertinent covariances.

The linear growth curve model is deterministic within an individual: Knowing the level and slope at time $t-1$ provides complete information on the true underlying status of a system at time. In other words, the intercept and growth rate are fixed within an individual in conventional linear growth curve models, although each individual is allowed to have his/her own intercept and growth rate. This is a key feature that distinguishes the linear growth curve model from the other two models considered here, the LLT model and the RW model: The variances associated with the intercepts and slopes in Σ_0 represent *between—person differences* in intercepts and slopes, rather than *within—person* deviations in the two components *over time*. Thus, in the context of the SF episode, the infant is expected to show a linear decrease in valence at a fixed rate over time, although infants in different dyads can differ in their average valence decline rate. In contrast, in the LLT model, interindividual differences in change are manifested through the individualized local levels and local slopes.

We took a group-based approach wherein modeling parameters were constrained to be invariant across all dyads in the sample. All three variations of the LLT model were fitted to the FFSF data. Maximum likelihood estimates of all of the freed parameters were obtained by optimizing the so-called prediction error decomposition function, a raw data loglikelihood function that can be computed using by-products from running the Kalman filter. The fixed interval smoother (FIS) was then used to obtain smoothed trend estimates (for estimation details see Durbin and Koopman 2001).

A plot of the data and the estimated trends of one randomly selected infant during the SF episode across all three models are shown in Fig. 3.2a–c. Results from fitting the LLT model indicated that the process noise variances for the local slopes were not significantly different from zero, suggesting that a RW model might serve as a more parsimonious model for the trends observed in the present data. To proceed with the RW model, we estimated the average initial levels in a and the corresponding variance elements in Σ_0, denoted as $\sigma^2_{\mu_{0,infant}}$ and $\sigma^2_{\mu_{0,mother}}$, as parameters in the model. Inspection of the resultant trend estimates in Fig. 3.2b compared to those in Fig. 3.2c confirmed our earlier conjecture: The trend estimates from the RW model were almost identical to those obtained using the LLT model, as expected due to the non-significant process noise variance estimates when the LLT model was fitted. Individual differences stemmed primarily from the influence of the process noise components on the local levels, yielding trends of very different shapes owing to the time-varying local levels.

A few other details from fitting the RW model should be noted. First, statistically significant interindividual differences were present in the initial valence of the mother ($\hat{\sigma}^2_{\mu 0,mother} = 0.88$, SE = 0.12), as well as the infant ($\hat{\sigma}^2_{\mu 0,infant} = 0.69$, SE = 0.10). However, the average initial levels were not significantly different from zero for both dyad members. Significant within-person deviations in the form of process noise were also found, with Var ($\zeta_{\mu_{infant,it}}$) and Var ($\zeta_{\mu_{mother,it}}$) both estimated to be significantly different from zero, with point estimates 0.08 and 0.17 and SEs 0.002 and 0.004, respectively.

The linear growth curve model was not very effective at capturing the shifts in level in this infant's data (see panel c in Fig. 3.2). When this model was fitted, the average intercept and slope parameters were not statistically different from zero, although significant inter-individual differences were found in the average intercepts for both dyad members (Var ($\mu_{intercept,infant}$) = 0.08, SE = 0.04; Var ($\mu_{intercept,mother}$) = 0.22, SE = 0.03). Allowing the average intercept and slope parameters to vary by FFSF episode did not change the results substantially. All three FFSF episodes were found to show average intercepts and slopes that were not significantly different from zero for both dyad members. The exception was an unexpected but small positive average slope (0.002, *SE* = 0.001) for mother during the SF episode, possibly reflecting the mothers' relief nearing the completion of the SF episode and the inadequacy of the deterministic linear growth curve model in capturing deviations from a perfect linear trajectory. No statistically significant inter-individual differences were found in the slopes. The deterministic nature of the model is evident in the smoothness of the trend trajectories in Fig. 3.2c.

Combining Models for Systematic Trends and Models for Intraidividual Variability

Results from the previous example indicated that among the three models considered, the bivariate RW model provided a better approximation for the overall upward/downward trends evidenced in the data. We now add other change components to the RW model to simultaneously represent the transient changes observed around the structured trends. Such models can be used to depict transient processes that may be indicative of an individual's ability to self-regulate or exhibit interactive influences with other family members. (Beebe et al. 2007; Chow et al. 2010; Kopp 1982; Schermerhorn et al. 2010; Stoller and Field 1982).

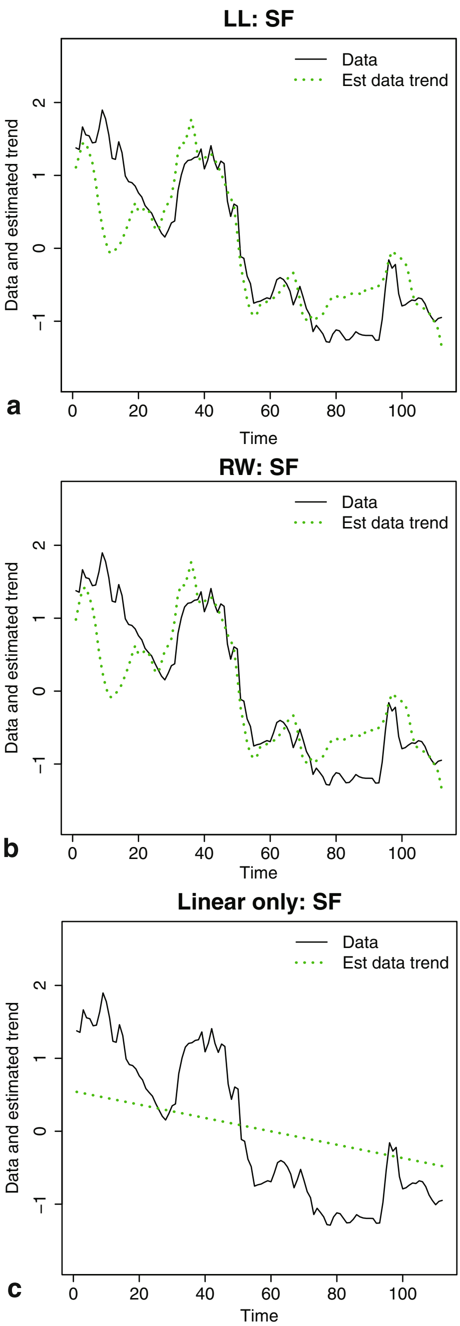


Fig. 3.2 A plot of the data and the estimated trends of one randomly selected participant obtained using: (**a**) the LLT model, (**b**) the RW model and (**c**) the linear growth curve model

Vector autoregressive (VAR) models are common choices for representing variability in the form of state fluctuations (e.g., Browne 1993; Browne and Nesselroade 2005). In the present context, the RW model can be combined with a VAR(2) model as

$$\begin{bmatrix} \mu_{infant,it} \\ \mu_{mother,it} \\ \alpha_{infant,it} \\ \alpha_{mother,it} \\ \alpha_{infant,i,t-1} \\ \alpha_{mother,i,t-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \phi_{i\rightarrow i,1} & \phi_{m\rightarrow i,1} & \phi_{i\rightarrow i,2} & \phi_{m\rightarrow i,2} \\ 0 & 0 & \phi_{i\rightarrow m,1} & \phi_{m\rightarrow m,1} & \phi_{i\rightarrow m,2} & \phi_{m\rightarrow m,2} \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} \mu_{infant,i,t-1} \\ \mu_{mother,i,t-1} \\ \alpha_{infant,i,t-1} \\ \alpha_{mother,i,t-1} \\ \alpha_{infant,i,t-2} \\ \alpha_{mother,i,t-2} \end{bmatrix} + \begin{bmatrix} \zeta_{\mu_{infant},it} \\ \zeta_{\mu_{mother},it} \\ \zeta_{\alpha_{infant},it} \\ \zeta_{\alpha_{mother},it} \\ 0 \\ 0 \end{bmatrix}, \tag{3.6}$$

where $\alpha_{j,it}$ represents member j in dyad i 's deviation from the trend at time t that is hypothesized to follow a VAR(2) process with influences both from the dyad member's own deviations as well as the other dyad member's deviations from two previous time points; $\phi_{j\rightarrow k,s}$ represents either an auto- or cross-regression parameter carrying the effect of member j ($m =$ mother; $i =$ infant) on member k from s time points ago. Autoregression effects capture the influence of a dyad member on him/herself from previous occasions. The occasion immediately preceding the current one is referred to as the lag-1 autoregression effect. The occasion prior to that is the lag-2 effect, and so on. Researchers have contended that such within-person influences represent a *self-regulation* attribute (Beebe et al. 2007; Chow et al. 2010). In contrast, the cross-member (i.e., the cross-regression) parameters represent the extent to which one dyad member is coupled to the affective valence of the other dyad member from previous time points and can be regarded as *interactive influence* parameters. Communicative dynamics within the various FFSF episodes can be conceived as a process through which the discrepancies between two partners' emotional states are minimized. Larsen (2000) used the analogy of a "thermostat" to describe such a dynamic process. For instance, immediately after the SF episode, the mother is likely to occupy a more positive set-point than the infant. In eliciting more positive responses from the infant, the mother is acting to shift the infant to a more positive set-point. The mother's effectiveness in bringing the infant closer to her (hypothetically more positive) set-point depends on the extent to which the cross-regression influence from mother to infant dominates over the influence from infant to mother. The speed with which the dyad is able to reestablish the necessary interactive connections to facilitate the return to more desirable set-points following a perturbation such as the SF manipulation can be interpreted as a resilience mechanism (Chow et al. 2005).

The corresponding measurement model is expressed as

$$\begin{bmatrix} Infant_{1,it} \\ Infant_{2,it} \\ Infant_{3,it} \\ Mother_{1,it} \\ Mother_{2,it} \\ Mother_{3,it} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ \lambda_{21} & 0 & \lambda_{21} & 0 & 0 & 0 \\ \lambda_{31} & 0 & \lambda_{31} & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & \lambda_{52} & 0 & \lambda_{52} & 0 & 0 \\ 0 & \lambda_{62} & 0 & \lambda_{62} & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} \mu_{infant,it} \\ \mu_{mother,it} \\ \alpha_{infant,it} \\ \alpha_{mother,it} \\ \alpha_{infant,i,t-1} \\ \alpha_{mother,i,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{infant1,it} \\ \varepsilon_{infant2,it} \\ \varepsilon_{infant3,it} \\ \varepsilon_{mother1,it} \\ \varepsilon_{mother2,it} \\ \varepsilon_{mother3,it} \end{bmatrix}, \tag{3.7}$$

Due to the complexity of the model, we simplified the initial condition specification based on earlier results from fitting the RW model and set the means of all initial latent variables to zeros (i.e., all elements in α are set to zero). In addition, we estimated the variances of the initial local levels of both infant and mother but fixed the variances of other initial VAR components to zero in Σ_0. A path diagram representation of the bivariate RW model with VAR(2) process is shown in the top panel of Fig. 3.3.

It is reasonable to expect the regulatory and interactive dynamics of the dyads to differ by FFSF episode. To test this, we allowed all statistically significant auto- and cross-regression parameters to vary by episode. The FF episode was used as the baseline condition, and deviations in auto- and cross-regression effects relative to the FF episode were estimated for the SF and RE conditions. Here, we focus on elaborating the results from a model wherein only the statistically significant parameters were retained. The parameter estimates from this model are summarized in Table 3.1. First, we note that consistent with the modeling results obtained when the RW only model was fitted, significant between-dyad differences were found in the average intercepts for both dyad members, indicating that the dyads started out with substantially different valence levels at the first observed time point.

The statistically significant lag-1 and lag-2 autoregression parameters (including $\phi_{i\to i,1}$, $\phi_{i\to i,2}$, $\phi_{m\to m,1}$, and $\phi_{m\to m,2}$) suggested that infants and mothers both show continuity in how they regulate the deviations in their valence levels around the trends that emerged with the FFSF manipulation. That is, the trends that emerged during each of the FFSF episodes (e.g. downward decline in the valence of infant during the SF, and upward increase in valence of infant as the mother attempts to calm and coax the infant during RE), as captured by the RW model, constituted the set-points in the RW plus VAR(2) model. The statistical significance of the lag-1 cross-regression parameters (i.e., $\phi_{i\to m,1}$ and $\phi_{m\to i,1}$) suggested that mothers and infants show bidirectional interactive influence as such regulatory actions unfold. Surprisingly, no significant difference was observed in the auto- or cross-regression parameters during the SF when compared to the FF episode. The only significant between-episode difference resided in the slight increase in the mother's lag-2 autoregressive parameter during the RE compared to the FF episode.

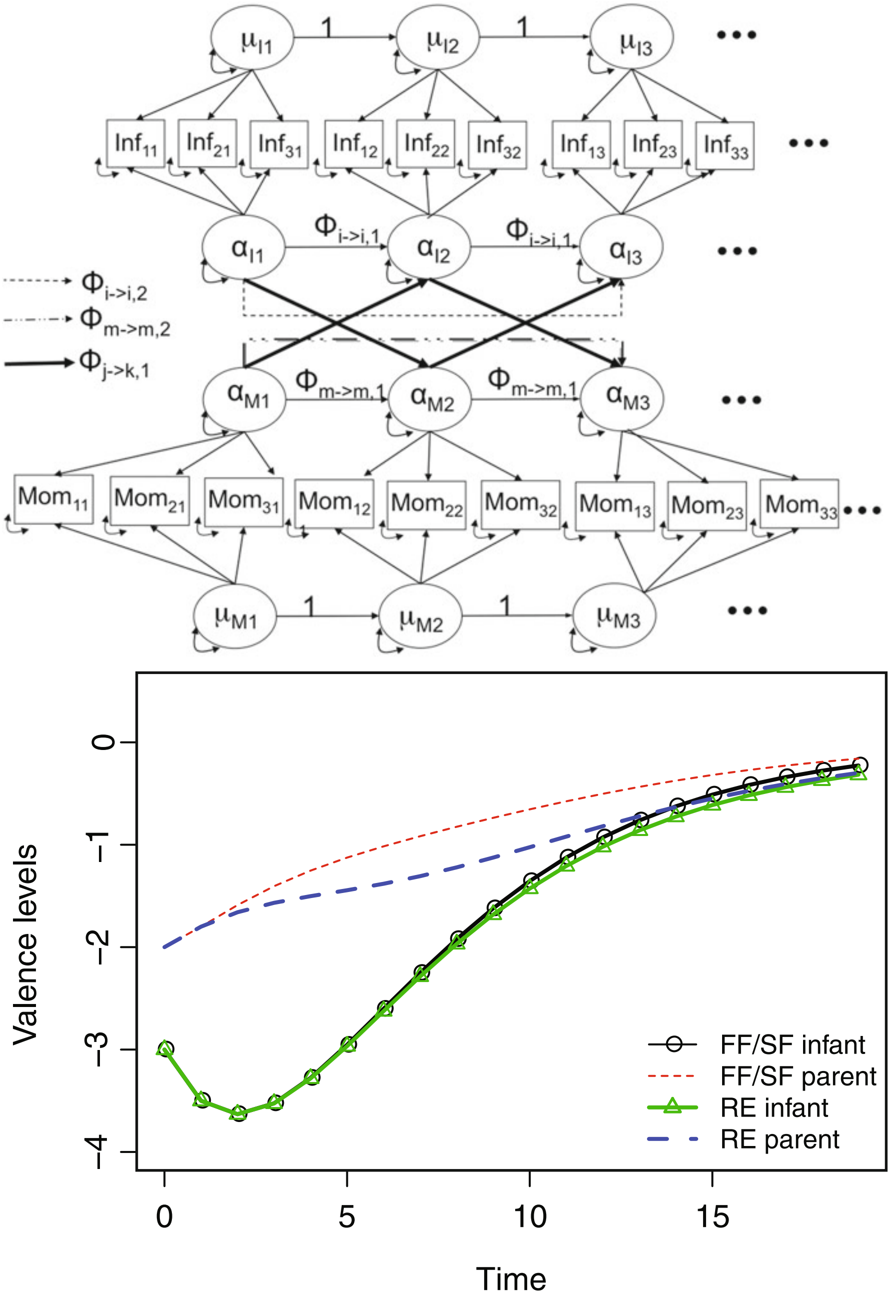


Fig. 3.3 (*Top panel*): A path diagram representation of the RW model with VAR(2) processes. $Inf_{ht} = h$ th observed item parcel for infant at time t, $Mom_{ht} = h$ th observed item parcel for mother at time t, $\phi_{i \to k,s}$ = auto- or cross-regression parameter from dyad member j to dyad member k from the s th previous occasion, μ_{It} = local level for infant at time t, μ_{Mt} = local level for mother, α_{It} = VAR process for infant at time t, α_{Mt} = VAR process for mother at time t. (*Bottom panel*): Hypothetical recovery trajectories simulated using parameters from the RW plus VAR(2) model with only the statistically significant parameters retained. The set-points for both infant and mother were set to be zero in this illustration

Table 3.1 Parameter estimates for the random walk with VAR(2) process model

Labels	Parameters	Estimates (SEs)
Factor loadings		
	λ_{21}	0.94 (0.01)
	λ_{31}	1.01 (0.01)
	λ_{52}	0.92 (0.01)
	λ_{62}	1.58 (0.02)
Process noise variances		
Local level, infant	Var $(\zeta_{\mu_{infant},it})$	0.004 (0.001)
Local level, mother	Var $(\zeta_{\mu_{mother},it})$	0.005 (0.001)
VAR(2), infant	Var $(\zeta_{\alpha_{infant},it})$	0.03 (0.001)
VAR(2), mother	Var $(\zeta_{\alpha_{mother},it})$	0.04 (0.001)
Measurement error variances		
	Var $(\varepsilon_{infant1,it})$	0.33 (0.01)
	Var $(\varepsilon_{infant2,it})$	0.24 (0.01)
	Var $(\varepsilon_{infant3,it})$	0.21 (0.01)
	Var $(\varepsilon_{mother1,it})$	0.67 (0.01)
	Var $(\varepsilon_{mother2,it})$	0.62 (0.01)
	Var $(\varepsilon_{mother3,it})$	0.06 (0.002)
Initial condition variances		
	Var $(\mu_{intercept,infant})$	0.52 (0.08)
	Var $(\mu_{intercept,mother})$	0.39 (0.06)
VAR parameters		
Lag-1 autoregression, infant	$\phi_{i\rightarrow i,1}$	1.55 (0.01)
Lag-2 autoregression, infant	$\phi_{i\rightarrow i,2}$	−0.61 (0.01)
Lag-1 autoregression, mother	$\phi_{m\rightarrow m,1}$	1.28 (0.02)
Lag-2 autoregression, mother	$\phi_{m\rightarrow m,2}$	−0.41 (0.02)
Lag-1 cross-regression, mother → *infant*	$\phi_{m\rightarrow i,1}$	0.02 (0.004)
Lag-1 cross-regression, infant → *mother*	$\phi_{m\rightarrow i,1}$	0.03(0.003)
Deviation in mother lag-2 autoregression during RE	$\Delta\ \phi_{m\rightarrow m,2,RE}$	0.02 (0.009)

It is important to emphasize that the auto- and cross-regression parameters function together to determine if and how the dyads are able return to their set-points. To help shed light on the collective effects of all the auto- and cross-regression parameters at different lags, we simulated the self-regulatory trajectories that are expected to emerge in the different FFSF episodes based on the parameter estimates from model fitting and plotted these trajectories (see bottom panel of Fig. 3.3). The set-points were constrained to be zero in this illustration to ease presentation. The trajectories illustrate a hypothetical scenario in which both infant and mother started out with low initial valence at $t = 0$. In the absence of further shocks, the trajectories help show the recovery pathways of the two dyad members. The lag-1 and lag-2 autoregression parameters, in this range, produced slight oscillatory patterns in the mother's and the infant's deviations in valence levels as these deviations were successively minimized toward the set-point. The small but positive lag-1 cross-regression parameters led to slight delays in both the mother's as well as the infant's recovery trajectories. The FFSF inherently creates an environment where mothers and infants are characterized by very different goals or set-points. While the infants' tendency is to gravitate toward a lower set-point (as a result e.g., of the SF manipulation), the mothers' goal is to bring the infants to a higher set-point. Thus, the cross-regression parameters suggested that mothers and infants tend to exert an antagonistic influence on each other in delaying the other party to return to their set points. The increase in lag-2 autoregression for mother during the RE episode resulted in slightly longer recovery time for the mother to return to her set-point in Fig. 3.3. Even though the infant's dynamic parameters remained unaltered compared to the FF episode, a slightly slower recovery trajectory was also observed for the infant as a result of the coupling dynamics between the two dyad members. This reflects the added struggles mothers are faced with during the RE in regulating both their own as well as the infants' valence levels.

Discussion

The state–space modeling techniques considered expand the work of many others who have used both exploratory as well as confirmatory methods to represent change (e.g., Ram, Chap. 2). Other structured nonlinear approaches (e.g., Browne 1993) or nonlinear mixed effects models (Davidian and Giltinan 1995) typically incorporate nonlinear trends while imposing some sort of linearization constraints in the estimation process while accommodating other serial dependency in the within–person residuals using e.g., an ARMA model (Verbeke and Molenberghs 2000). The immediate appeal of the RW model and the more complex LLT model considered in the present article is that it is linear, and yet it is flexible enough to capture a variety of non–systematic trends. This is a much needed feature for modeling processes that do not necessarily show trends that can be readily described using parametric functions (e.g., affective changes during dyadic interaction) and this chapter is one of the first illustrations of how such non-systematic trends can be captured and combined with models of intraindividual variability. On the contrary, one major strength of taking a parametric approach to modeling nonlinear trends is that the parameters from the models are usually of direct theoretical interest and are readily interpretable to the researchers. Thus, these models are important from a modeling standpoint. However, if trends are not the focus of interest, detrending procedures are likely needed to avoid introducing biases into other intraindividual variability parameters (Craigmile et al. 2009).

We took a group-based intensive longitudinal data approach to constructing our illustrative examples. The assumptions imposed are similar to those typically assumed in fitting group-based panel models. Statistical conditions that need to be fulfilled before one can justify "pooling" a group of individuals deserve further investigation (see e.g., Molenaar 2004). In cases where a researcher wishes to pursue model fitting at the individual level, the modeling approach shown here can be readily used.

Devising models for intraindividual variability requires researchers to formulate research hypotheses pertaining to change from a within–individual perspective. Research accumulated in the past two decades has led to fruitful conclusions concerning the importance of studying the *magnitudes* of intraindividual variability across a broad array of constructs. Very little is known, however, concerning the *structures* and *time course* of intraindividual variability as a process. Fortunately, social and behavioral scientists are better equipped now than ever before to advance premises pertaining to intraindividual variability. However, some techniques that are well suited for analyzing the data that arise from intraindividual variability studies, including the approaches presented here, have not been widely utilized. We believe that the illustrative examples provided here can encourage more researchers to explore and develop other ways to study intraindividual variability, both from an intra- as well as an inter-individual perspective.

References

Adamson, L. B., & Frick, J. E. (2003). The still face: A history of a shared experimental paradigm. *Infancy, 4*, 451–473.

Allaire, J. C., & Marsiske, M. (2005). Intraindividual variability may not always indicate vulnerability in elders' cognitive performance. *Psychology and Aging*, *20*, 390–401.

Almeida, D. M., Piazza, J. R., & Stawski, R. S. (2009). Interindividual differences and intraindividual variability in the cortisol awakening response: An examination of age and gender. *Psychology and Aging*, *24*(4), 819–827.

Beebe, B., Jaffe, J., Buck, K., Chen, H., Cohen, P., Blatt, S., Kaminer, T., Feldstein, S., & Andrews, H. (2007). Six–week postpartum maternal self-criticism and dependency and 4–month mother-infant self- and interactive contigencies. *Developmental Psychology*, *43*(6):1360–1376.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, *54*, 579–616.

Bollen, K. A., & Curran, P. J. (2004). Autoregressive latent trajectory (ALT) models: A synthesis of two traditions. *Sociological Methods & Research, 32,* 336–383.

Browne, M. W. (1993). Structured latent curve models. In C. M. Cuadras & C. R. Rao (Eds.), *Multivariate analysis: Future directions 2* (pp. 171–198). Amsterdam: North–Holland.

Browne, M. W., & Nesselroade, J. R. (2005). Representing psychological processes with dynamic factor models: Some promising uses and extensions of autoregressive moving average time series models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A Festschrift for Roderick P. McDonald* (pp. 415–452). Mahwah: Erlbaum.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, *101*, 147–158.

Chow, S. -M., Haltigan, J. D., & Messinger, D. S. (2010). Dynamic patterns of infant-parent interactions during face-to-face and still-face episodes. *Emotion*, *10*(1), 101–114.

Chow, S. -M., Ram, N., Boker, S. M., Fujita, F., & Clore, G. (2005). Emotion as thermostat: Representing emotion regulation using a damped oscillator model. *Emotion, 5*(2), 208–225.

Chow, S. -M., Ho, M. -H. R., Hamaker, E. J., & Dolan, C. V. (2010b). Equivalences and differences between structural equation and state–space modeling frameworks. *Structural Equation Modeling, 17,* 303–332.

Collins, J. J., & De Luca, C. J. (1994). Random walking during quiet standing. *Physical Review Letters*, *73*(5), 764–767.

Craigmile, P. F., Peruggia, M., & Van Zandt, T. (2009). Detrending 2 time series. In S. -M. Chow, E. Ferrer, & F. Hsieh. (Eds.). *Statistical methods for modeling human dynamics: An interdisciplinary dialogue* (pp. 213–240). Taylor & Francis, New York.

Davidian, M., & Giltinan, D. M. (1995). *Nonlinear models for repeated measurement data*. London: Chapman & Hall.

Doornik, J. A. (1998). *Object–oriented matrix programming using Ox 2.0*. Timberlake Consultants Press, London.

Durbin, J., & Koopman, J. S. (2001). *Time series analysis by state space methods*. New York: Oxford University Press.

Eid, M., & Diener, E. (1999). Intraindividual variabiliy in affect: Reliability, validity and personality correlates. *Journal of Personality and Social Psychology, 76*(4), 662–676.

Ekas, N., Haltigan, J., & Messinger, D. S. (2013). The dynamic still-face effect: Do infants decrease bidding over time when parents are not responsive? *Developmental Psychology, 49*(6), 1027–1035.

Frazier-Wood, A., Bralten, J., Arias-Vasquez, A., Luman, M., Ooterlaan, J., Sergeant, J., Rommelse, N. N. (2012). Neuropsychological intra-individual variability explains unique genetic variance of ADHD and shows suggestive linkage to chromosomes. *American Journal of Medical Genetics B: Neuropsychiatric Genetic, 159B*(2), 131–140.

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2002). On the nature of SEM estimates of ARMA parameters. *Structural Equation Modeling*, *9*(3), 347–368.

Harvey, A. C., & Souza, R. C. (1987). Assessing and modelling the cyclical behaviour of rainfall in northeast Brazil. *Journal of Climate and Applied Meteorology, 26*, 1317–1322.

Kishton, J. M., & Widaman, K. F. (1994). Unidimensional versus domain representative parceling of questionnaire items: An empirical example. *Educational and Psychological Measurement*, *54*, 757–765.

Koopman, S. J., Shephard, N., & Doornik, J. A. (1999). Statistical algorithms for models in state space using ssfpack 2.2. *Econometrics Journal*, *2*, 113–166.

Kopp, C. B. (1982). Antecedents of self-regulation: A developmental perspective. *Developmental Psychology*, *18*, 199–214.

Kurdek, L. A. (2005). Gender and marital satisfaction early in marriage: A growth curve approach. *Journal of Marriage and Family, 67*(1), 68–84. doi:10.1111/j.0022-2445.2005.00006.x.

Larsen, R. J. (2000). Toward a science of mood regulation. *Psychological Inquiry*, *11*, 129–141.

McArdle, J. J., & Nesselroade, J. R. (2003). Growth curve analysis in developmental research. In J. Schinka & W. F. Velicer (Eds.). *Handbook of psychology: Volume 2, Research methods in psychology, pages* (pp. 447–480). New York: Pergamon Press.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107–122.

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific pyschology–this time forever. *Measurement: Interdisciplinary Research and Perspectives, 2*, 201–218. doi:10.1207/s15366359mea0204_1.

Nesselroade, J. R., & Ford, D. H. (1985). P–technique comes of age: Multivariate, replicated, single–subject designs for research on older adults. *Research on Aging*, *7*, 46–80.

Schermerhorn, A. C., Chow, S. -M., & Cummings, M. E. (2010). Dynamics of family influence processes during interparental conflict. *Child Development*, *46*(4), 869–885.

Sliwinski, M. J., Almeida, D. M., Smyth, J., & Stawski, R. S. (2009). Intraindividual change and variability in daily stress processes: Findings from two measurement-burst diary studies. *Psychology and Aging*, *24*(4), 848–840.

Stoller, S., & Field, T. (1982). Alteration of mother and infant behavior and heart rate during a still-face perturbation of face-to-face interaction. In T. Field & A. Fogel (Eds.). *Emotion and early interaction* (pp. 57–82). Erlbaum, Hillsdale.

Tronick, E., Als, H., Adamson, L., Wise, S., & Brazelton, T. B. (1978). The infant's response to entrapment between contradictory messages in face-to-face interaction. *Journal of the American Academy of Child Psychiatry*, *17*, 1–13.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. Springer–Verlag, New York.

Walls, T. H., & Schafer, J. L. (2006). *Models for intensive longitudinal data*. Oxford: University Press.

Chapter 4
The Benefits and Challenges of Modeling Intra-Family Variability

Andrew J. Fuligni

Families are dynamic entities that fluctuate across days, weeks, and years. Key aspects of family interactions, such as conflict and social support, may differ across time and situations. Unfortunately, our traditional methods generally focus on mean levels of family processes, and therefore under-characterize the complexity of families. The chapters authored by Teachman (Chap. 1), Ram, Shiyko, Lunkenheimer, Doerksen, and Conroy (Chap. 2), and Chow, Messinger, and Mattson (Chap. 3) provide unique ways to model family variability as measured in repeated-measures and multiple family member designs. As described below, these approaches offer both benefits and challenges to the family researcher interested in family variability. In this commentary of Chapters 1-3, I first describe examples of the unique understanding of families that is provided by modeling intra-family variability. I then discuss the challenges, in terms of the additional data, time, and resources that are needed to conduct such studies. Finally, I close with a demonstration of the benefits and challenges by applying the models and analytical techniques offered by the authors to hypothetical studies of intra-family variability in sleeping behavior.

The Benefits of Modeling Intra-Family Variability

Repeated measures of family experiences provide unique information that can improve the precision and richness of our models of the family. For example, intensive measurements of family interactions and events across the days or weeks can inform estimates of family routines and the predictability of family experiences. We can also understand how families respond to external shocks such as the loss of a job or an economic downturn, and gain information about how experiences outside the family, such as at work or school, spillover and effect internal family dynamics and multiple family members. Repeatedly assessing interactions or experiences can tell us how individual family members influence one another across time, such as through

A. J. Fuligni (✉)
Department of Psychiatry & Biobehavioral Science, University of California,
Los Angeles, CA, USA
e-mail: afuligni@ucla.edu

S. M. McHale et al. (eds.), *Emerging Methods in Family Research,*
National Symposium on Family Issues 4, DOI 10.1007/978-3-319-01562-0_4,

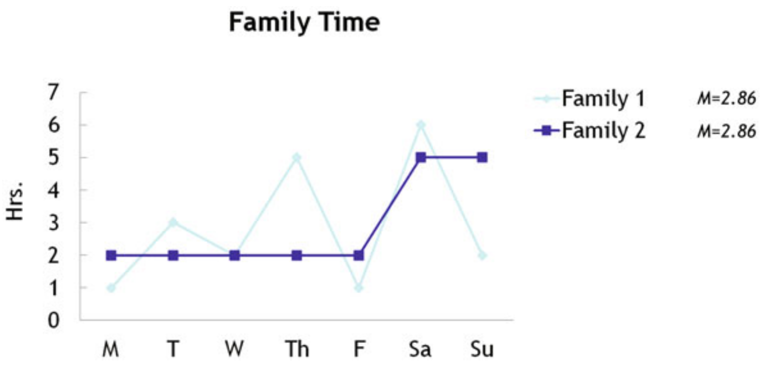


Fig. 4.1 Estimating family routines and predictability

emotional transition or the enactment of social support. Longitudinal studies allow for estimation of the impact of developmental change, such as puberty or aging, on how family members interact and exchange resources with one another. Finally, intensive repeated measures allow us to treat individuals and families as their own controls in non-experimental designs. By accounting for the innumerable ways that families and individuals differ from one another, such fixed-effect models allow us to have more confidence that the associations we estimate are attributable to the constructs of interest, as opposed to unobserved individual and family variations.

The value of measuring routines and predictability is demonstrated in Fig. 4.1. The two families represented in the graph spend the same average amount of time together across the week. A one-time survey of these families asking about the typical amount of family time would suggest that these two families are the same. Intensive repeated measures across the days of the week, however, demonstrate how family time is distributed quite differently across the two families. Time together in Family 1 is unpredictable across the days, consisting of wide swings from day to day without following a pattern. In contrast, time together in Family 2 is quite predictable. On weekdays, Family 2 consistently spends only about 2 h together. In comparison, the weekends appear to be days of dedicated family time such that the family spends more than twice the amount of time together on Saturdays and Sundays as they do on weekdays. It is possible that the difference between these two families in the predictability of their family time across the week produces differences in the adjustment of individual family members beyond the effect of the average amount of time that they spend together.

Figure 4.2 shows the benefit of documenting information about family reactivity to external shocks and stressors. Partners in Families 1 and 2 average the same frequency of conflict across the days of the week. A design that measures dyadic conflicts and stressors experienced by one of the partners across the week, however, shows differences between the two families in the extent to which stressors at work for one partner spill over into the interactions with the other partner. Across the first four days of the week, Family 1 experiences seemingly higher levels of partner conflict

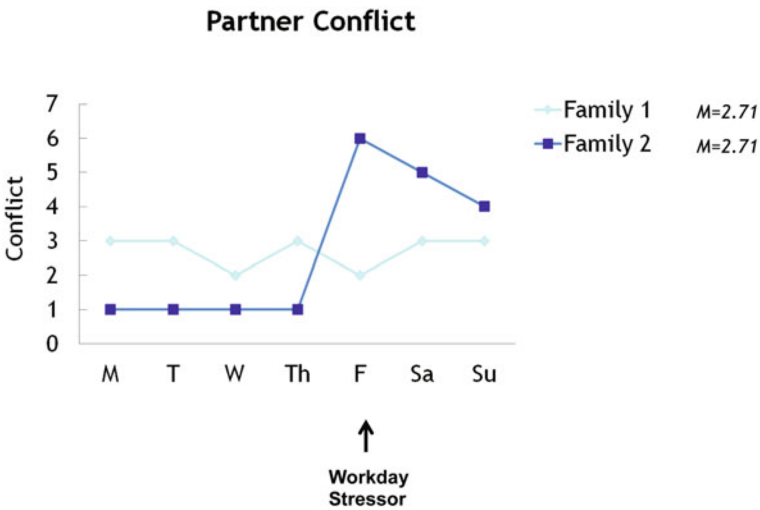


Fig. 4.2 Estimating reactivity and spillover

than Family 2. The partners' relationship in Family 2 appears to be quite harmonious across the first four days of the week, until one of the partners experiences a work stressor on Friday. On that day, the level of conflict spikes dramatically to twice the level that the partners experienced earlier in the week, and conflict remains high even 1 and 2 days after the experience of the workday stressor. Conflict between partners in Family 1, in contrast, is fairly stable across the days and does not differ as a function of stressors experienced at the workplace. Such a notable difference between these two families would not have been observed without the intensive repeated measurement of this design.

Research designs that capture family fluctuation also tell us about how family members influence one another across time. Figure 4.3 represents the level of happiness experienced by the father and the mother within the same family. The occurrence of a moody teenager is shown below the X axis. As Fig. 4.3 demonstrates, the father and mother report the same average level of happiness across the week, but they differ dramatically in the extent to which their happiness varies as a function of whether or not their teenager is particularly moody that day. The father's happiness appears fairly stable and largely independent of the mood of his teenager. The mother's mood, in contrast, is conditional upon the mood of her teenager and as shown in the figure, the mother's happiness drops considerably on days in which her teenager is moody. As a result of this dependence, the happiness of the mother fluctuates widely across the days, similar to the spillover results demonstrated in Fig. 4.2. The results shown in Fig. 4.3 demonstrate the complexity by which individual family members can affect one another in ways that are not observable in traditional single measurement designs.

Of critical importance to the study of families is the question of how developmental changes in family members influence family process. Keeping with the example of

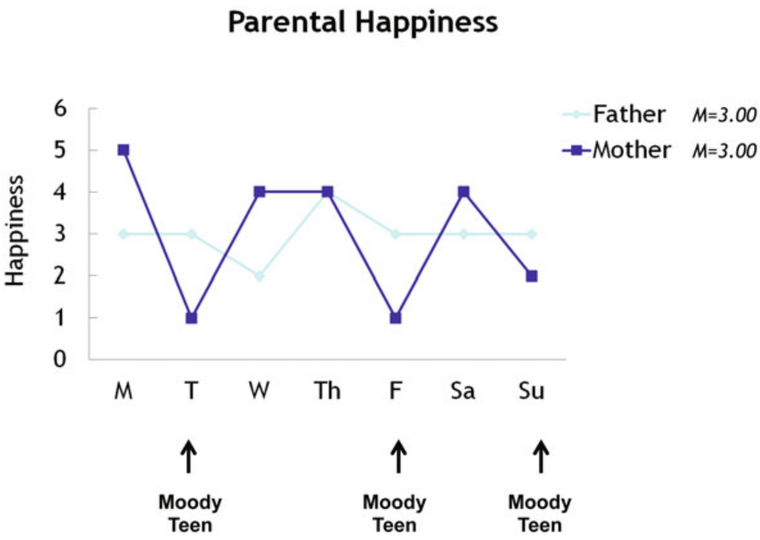


Fig. 4.3 Estimating family members' influence on each other

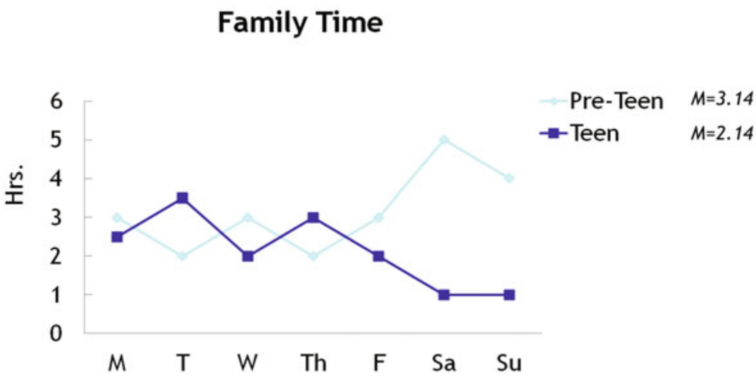


Fig. 4.4 Estimating the impact of developmental change

the repeated measures design across days of the week, Fig. 4.4 shows how time spent with the family differs according to the developmental stage of the child. On average, the preteen spends more time with the family than the teenager. Detailed assessment across the week, however, reveals how this developmental difference is largely a weekend difference. During the week on school days, these two individuals spend about the same amount of time with the family. Weekends, however, are a very different story. The preteen increases the amount of time spent with the family on weekends perhaps due to more family activities or simply more time spent at home. In contrast, the teenager spends dramatically less time with the family on Saturday and Sunday. Perhaps because of increased time spent with friends or more frequent

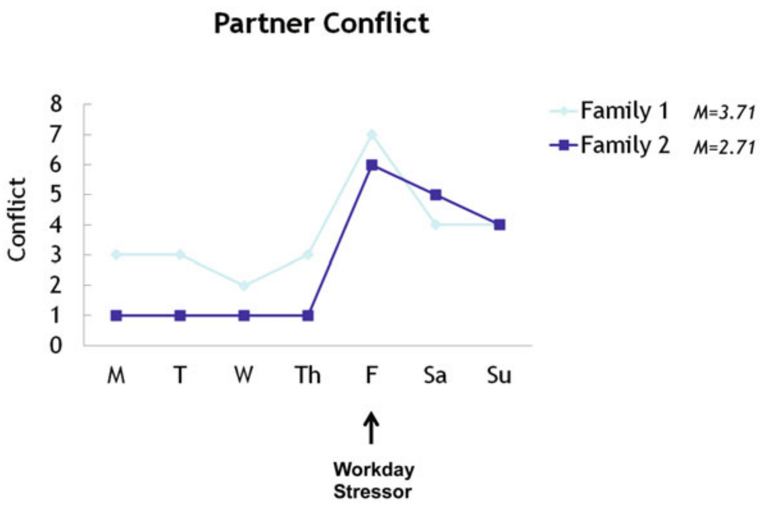


Fig. 4.5 Using families as their own controls

extracurricular activities that are available during the high school years, the teenager spends very little time with the family on the weekends.

Finally, Fig. 4.5 displays perhaps one the most valuable features of an intensive repeated measures design. Family 1 and Family 2 average different levels of partner conflict. Although not shown, these two families also differ in the same direction in the frequency of work-day stressors experienced by one of the partners, and a significant correlation between these two means shows that family differences in workplace stressors are associated with family differences in partner conflict. The problem is that there may be additional, unmeasured differences between the two families that could account for this correlation. As Fig. 4.5 shows, a within-family design across the days of a week suggests that workday stressors are indeed associated with changes in partner conflict in both families. By examining this process within families across time, we essentially control for unmeasured differences between the two families and therefore have more confidence that the observed association is primarily due to the constructs being examined.

The Challenges of Modeling Intra-Family Variability

The three chapters by Teachman, Ram, Chow, and colleagues take advantage of the value of measuring intra-family variability and offer new ways to model the complex types of data obtained by such designs. As a collection, the three chapters make a nice contribution in that they model intra-family variability at multiple levels. Teachman (Chap. 1) focuses on change and variability across years and development. Ram et al. (Chap. 2) examine daily variability in family interactions, and Chow et al. (Chap. 3) take a micro-analytic approach to examine momentary change and variability in the interactions between family members.

Although the authors focus on statistical modeling, the chapters make substantive contributions that go beyond the specific questions addressed in their empirical examples. Teachman's (Chap. 1) modeling of both fixed and random effects in longitudinal data demonstrates how such an approach can assist in the identification of when and where group differences (e.g., ethnic variability) in key measures of health and adjustment (e.g., obesity) emerge across time. Teachman's explication of a paired-process model shows how family researchers can address the question of how family members contribute to change in each other over time. Ram et al.'s (Chap. 2) modeling of the mood of new parents before and after the birth of their child offers a nice way to examine how the self-organizing and self-sustaining dynamics within families change as a function of the addition of a newborn infant to the family. Finally, Chow et al. (Chap. 3) argue that it is not enough just to examine family variability across time. They offer analytical techniques to address the structure and time-course of the intra-family variability, such as the initiation and rate of change across time.

The benefits of intra-family designs, however, do not come easily or cheaply. As evident from the three chapters and the examples provided earlier in this chapter, intra-family designs have substantial data demands. The modeling of longitudinal data described by Teachman requires the tracking, monitoring, and retention of a large sample of individuals and families over several years. The data used by Ram and colleagues were obtained from a daily diary checklist study that took place over a period of several weeks. Chow and colleagues' micro-analysis of parents and infants required the measurement of face-to-face interactions of the dyads across different conditions of the experiment. In addition to collecting such data, Chow et al.'s analysis required the coding of facial expressions of emotion by multiple, independent coders. Finally, the examples that I provided require the collection of data on a daily basis from multiple family members. These types of data are labor intensive to collect, place extra demands on the participants themselves, and require extensive data management and processing even before the analyses can take place. In addition to the time and effort of both the scientists and the participants, these types of intensive repeated measurements can become quite expensive.

Because of the demands of data collection, both in terms of time and sample retention, the designs discussed in chaps. 1, 2, and 3 generally involve smaller and by extension, more selective samples. Requiring multiple members of the family to participate, as in the studies presented by the authors, inevitably creates a sample that is somewhat biased toward those who are more willing to participate in a study. The experimental data analyzed by Chow et al., (Chap. 3) in particular, requires caregivers who have the time, flexibility, and interest to travel to a research laboratory. The data analyzed by Teachman (Chap. 1) were more representative by virtue of the sampling strategies in the two studies, but the sample sizes differed as a function of the demandingness of the study. The larger, more representative study included only one individual from a family, whereas the couples study had a significantly smaller sample. Comments about the size and representativeness of the samples in these chapters are not criticisms of the studies themselves; rather, they are meant to highlight the impact of intensive repeated measurements on sample characteristics.

As noted by Chow and colleagues, an additional challenge of intensive, intra-family models is that we often have limited theory to guide the design, collection, and analysis of these data. For example, there is little basis for determining the appropriate time interval between repeated measurements for many studies. Although yearly data collection may make intuitive sense for many developmental designs, it is unclear whether 12 months represents the actual time it takes for parents to effect children or vice versa. Many daily diary studies employ data collection on a nightly basis so that participants can sum their experiences across the day, but it is possible that the causal action between key constructs occurs in a matter of hours or minutes. When data collection occurs continuously across a period of time, such as in the videotaping of family interactions, decisions still need to be made as to the length of the time intervals that are coded and analyzed. The ideal solution, of course, would be to collect data on as short a time interval as possible and use comparative modeling to determine the actual time of the causal process. Aside from smaller studies that are able to employ a continuous data collection, this strategy unfortunately is not feasible, efficient, or reasonable in terms of cost. It is very challenging, therefore, to decide upon a data collection strategy that is both logistically feasible and theoretically meaningful.

Finally, it is important to remember that conclusions that focus on the variability of the processes being measured may be confounded with the design of the study. Greater variability will be evident in designs that employ more frequent measurements that are closer in time. This will affect conclusions about variability in the constructs of interest and the ability to model associations between two family processes. For example, the parallel process model estimated by Teachman might have produced even stronger associations between the happiness of partners if they were measured more frequently across time. The brief dip in the emotional valence of the infants' facial expressions after the experimental manipulation reported by Chow et al. likely would not have been observed had the time interval of the measurements been longer. Although the confounding between observed variability and the design of the study seems to be an obvious point, it nevertheless should be kept in mind when the main questions of intra-family studies focus on resolving issues of variability vs. stability in family processes.

The challenges of conducting intra-family studies, therefore, leave lingering questions about the balance between the costs of such designs and their benefits. How do we determine the added value of such intra-family designs as compared to traditional designs? Clearly, intra-family designs and their intensive repeated measurements of the same family provide richer descriptions of family processes. As shown in Fig. 4.2, two families may differ dramatically from one another in terms of the distribution and variability in partner conflict across the week, despite having the same level of average daily conflict. But is better description enough of an added value to warrant the time and resources necessary to conduct such a daily level study? Similarly, both Teachman (Chap. 1) and Chow et al. (Chap. 3) conduct comparative estimates of model fits to determine which models best describe the observed data. But is model fit of observed data the best determination of the added value of a particular technique? Do these modeling techniques or designs provide a substantively different and more valuable understanding of family processes than is available with a more traditional research design? Do they change previous conclusions about the family experiences that are being studied? Do they better predict indicators of individual health and adjustment? Do they suggest alternative types of intervention and prevention? Some of these questions are tied to the basic science justification for research and others are linked to the public health justification. Therefore, the answer does not have to be yes to all of the questions, but the time, effort, and cost to conduct intra-family studies seem to necessitate addressing at least some of them.

The complexity of the modeling techniques and the specific nature of the substantive conclusions they provide raise additional questions about whether the value of the techniques will get lost in translation. For example, the processes studied by Ram et al. (Chap. 2) are rather difficult to explain to other family researchers. The principles the authors obtain from ecology are fascinating and seemingly appropriate for the study of families, thereby providing a potentially unique contribution to our understanding of family processes. Likewise, it is fairly easy to understand the basic principle that new parents are reacting to each other's emotions closely and that they work as a self-organizing system to try to maintain emotional balance in the home. But the modeling conducted by the authors is much more than that and potentially offers an understanding of the mechanisms by which this process occurs. It is possible, however, that these mechanisms will escape most family researchers who will instead see the contribution of the work at a more general level.

I have experienced this in my own work that employs within person, daily level analyses of adolescent experiences to control for unobserved individual differences. We do this in an attempt to gain a stronger inference about the potential impact of the experiences themselves on one another. For example, when discussing a recent study of ours on teenagers cramming the night before a test and the negative impact on their sleep time at the daily level (Gillen-O'Neel et al. 2013), media reports often described the results as suggesting that adolescents who studied more often tended to sleep less at night. It was difficult to communicate what we believed to be the true value of the study, which was the linking of these two activities within individual adolescents regardless of how much they studied or slept on average. Such problems of communication are not limited only to press accounts of research; I have seen similar mischaracterizations of our daily-level research in academic journals.

I do not mean to imply that the complexity of modeling intensive intra-family designs and the apparent lack of understanding about the complexity among other researchers means that such designs and statistical modeling should not be conducted. It is important for the research community to continue to stay educated on statistical and methodological advances. Nevertheless, the customer is always right, and it is critical for those who conduct such intra-family research and modeling to communicate the findings in a manner that is accessible while still capturing the unique contribution of the research. If this cannot be done successfully, then questions will continue to linger as to whether such efforts are worth the investment of time and resources. In summary, although intensive intra-family designs and modeling may be exciting for a certain type of family researcher, it remains imperative to address the question about when they are most useful and important relative to more traditional designs.

One Example: Sleep as a Family Process

In order to exemplify the balance between the benefits and challenges of intra-family designs, I apply the designs and modeling techniques offered by Teachman (Chap. 1), Ram et al. (Chap. 2), and Chow et al. (Chap. 3), to the study of sleep as a family process. My colleagues and I have become increasingly interested in the dynamics of teenage sleep, employing daily diary checklist methods to the analysis of daily variability in sleep in connection with variability in other experiences in adolescents' lives as well as aspects of their adjustment (Fuligni and Hardway 2006; Gillen-O'Neel et al. 2013) Up to this point, we have analyzed teenage sleep only as an individual activity. That is, we have measured and modeled only the time and frequency of adolescents' nightly sleep and its relation to other adolescent activities such as study time, socializing with friends, mood, and academic achievement. We now would like to try to contextualize adolescent sleep and study it in relation to the sleep of other family members. That is, we would like to study sleep as a family-wide practice, and not just as a behavior of single individuals. Although this has been done in prior research for new parents and their infants, little research has examined the sleep of multiple family members during the teenage years. Investigators increasingly have examined the role of family-level processes such as economic conditions, family climate, and parental monitoring of sleep (El-Sheikh et al. 2006; McHale et al. 2011; Meijer et al. 2001), but they have not yet done much work linking the sleep behaviors of different family members with one another.

In order to link the sleep behaviors of different family members with one another, I focus on the key question that the authors of chaps. 1, 2, and 3 addressed and try to apply their designs and techniques to a specific question about family sleep. I first discuss the question and the methods that could be used to address each question, followed by an assessment of the relative benefits and challenges of doing so.

How do Family Members Contribute to Change in Sleep Over Time? Several changes take place during the teenage years that impact adolescent sleep behavior. Puberty generates fundamental and biologically-driven changes in the circadian rhythm of children such that they increasingly prefer later bed and wake times (Carskadon 2002) For adolescents in the United States and some other countries, this delayed phase preference unfortunately is accompanied by increasingly earlier school start times during the middle and high school years (Carskadon et al. 2004). The collision between biological and institutional changes moving in different directions has been cited as one source of the sleep restriction that takes place during the teenage years (Dahl and Lewin 2002) Although this change has been studied frequently among teenagers, there has been little research examining how adolescents' sleep restrictions may influence the sleep of their parents. At the same time, there are substantial individual differences in teenage sleep behaviors such that some adolescents continue to obtain sufficient amounts of sleep during the teenage years. Whether these individual differences reflect family-wide sleep practices remains to be seen.

Teachman's (Chap. 1) parallel process model of happiness between married couples could be applied to the linkages between changes in the sleep practices of different family members over time. In this design, both teen and parent sleep is measured separately across three time points. Sleep could be estimated using multiple parameters, such as time, quality, bed time, and wake time. Such a model would allow us to estimate the correlation between parent and teen sleep at an early time point, perhaps before the beginning of puberty or the transition to secondary school, as well as whether the changes in one person's sleep is associated with changes in another person's sleep. As represented in Teachman's model, this approach does not allow for the estimation of the direction of the effect between parent and teen sleep, but it would be important to first establish whether sleep problems faced by the adolescent are shared by the parent. Additional specifications could be made to the model to estimate whether changes over time are attributable to pubertal changes in the teenager or institutional changes in their school start times, or even other factors such as school work load, or time spent socializing in the evening. Depending upon the interval between the measurements, cross-lagged models potentially could be employed to gain traction on the direction of the effect.

In terms of substantive contributions, such modeling of family sleep would be an important and unique addition. Theoretically, it would move thinking about teenage sleep to a consideration that sleep is just as much a family practice as it is an individual behavior. In our efforts to understand the variation in adolescent sleep behavior, therefore, we would need to address factors that may have an influence at the family level. At the same time, such modeling would enhance our understanding of parental sleep, with the potential suggestion that sleep problems during adulthood may be driven in part by the sleep problems of children. Although this reality has been documented during the years of infancy, there has been little attention to the possibility that this dynamic continues through the teenage years. Clinical and public health efforts to improve teenage and adults sleep, therefore, might profit from considering sleep as a family wide practice.

How Do Families Function as Self-Organizing and Sustaining Systems in Terms of Sleep? Ram et al. (Chap. 2) offer a unique way to examine how different members of the same family shape one another's emotional experience across time. Borrowing from the concepts of ecology, the authors propose a way to examine how new parents shape each other's emotional well-being before and after the arrival of their child. The basic principle is that there is a trend toward a homeostasis of emotion within the family, and their modeling strategy allows for the estimation of both the direction of that trend as well as how partners influence and react to one another over time.

Ram and colleagues' approach to examining systems of mutual influence within the family context could be used to examine how families function as self-organizing and sustaining systems in terms of sleep. Whereas the mood data collected by Ram and colleagues suggest that partners push and pull one another in different directions in order to maintain a homeostasis within the family, the process by which families function as self-organizing and sustaining systems in terms of sleep is likely to be different. In terms of sleep, it is probable that under normal conditions, family members will move each other in the same direction. That is, as one member of the family sleeps, the other is more likely to sleep as well. As one family member has difficulty sleeping, that difficulty is likely to affect another family member. Similarly, if there is a shock or disruption to the system such as an event that disrupts the sleep of one of the parents, the family's self-organizing tendency is likely to pull the other partner in the same direction in terms of sleep. Therefore, under normative conditions, the sleep patterns of one parent are likely to follow in the same direction as the other parent.

The birth of a child, however, could change the pattern dramatically. With a new infant in the household, sleep tends to be disrupted and more constrained. Therefore, the dynamics of the family system may change such that the premium is placed on at least one parent being able to sleep. A disruption to sleep, such the need to feed or change the infant, initially may wake both parents during the night. As a designated parent tends to the infant, however, the other parent is likely to go back to sleep. The organizing principle then becomes one in which parents take turns sleeping and caring for a waking infant. The analyses offered by Ram and colleagues might show that sleep of each individual partner goes in different directions when there is a shock to the system such as a waking infant who will not go back to sleep on its own.

Ram and colleagues suggest that their modeling techniques represent a way to assess the complex dynamics suggested by family systems theories. As such, their approach might offer a method to study whether sleep within the family operates according to the same principles as family systems theories. Such an empirical demonstration of the linkages between the sleeping behaviors of different family members would be a valuable addition to a field of study that has focused primarily on sleep as a property of the individual.

What is the Structure and Time Course of Intra-Family Variability in Sleep? It is somewhat more difficult to consider how the modeling strategies suggested by Chow et al. (Chap. 3) can be adapted to the study of family sleep. Nevertheless, just as their example focused on how parents try to coax emotional recovery and stability from their infant, this approach could be used to analyze how parents try to coax their infants to sleep.

Consider Fig. 3.3 (bottom panel) in Chap. 3. Their graphs represent the structure and time course of infant and parent emotional expression, starting at the beginning of a particular experimental condition. The same analysis and graphical representation could be used to examine the sleep levels of parents and infants after an awakening during the middle of the night due to a diaper change or a feeding. The parent initially is likely to be more awake than the infant, and becomes even more awake as he/she tends to the infant's need. The infant, as the need is being attended to, slowly begins to go back to sleep as the parents' sleep lags behind because he/she was more awake and remains awake until the infant is back to sleep. This representation of the process may be a normative pattern. Deviations from this pattern, such as an infant who becomes increasingly awake over time or a parent who takes much longer to return to sleep after getting the infant settled, could be indicators of sleep problems of individual members in the family that are nonetheless a function of family-wide processes.

Benefits and Challenges The approaches in Chaps. 1, 2, and 3 all offer interesting ways to examine the complexities by which individual family members influence the sleep of other family members. The three approaches demonstrate many of the benefits of intra-family designs. They nicely show how family sleep is not static and that focusing only on mean scores under-characterizes the nature of sleep processes within the family. The approach of Ram and colleagues, which treats families as self-sustaining and self-organizing systems, could be employed to capture sleep as a family routine and would allow for an analysis of the predictability of that routine. Interestingly, that predictability does not necessarily mean the same amount of sleep over time. Rather, predictability is the extent to which the sleep patterns up one family member influences the sleep of another family member in predictable ways.

Ram and colleagues' approach also focuses upon how family members react to shocks to the system. That shock may be the introduction of an infant to the family, which undoubtedly disrupts established patterns of sleep in the family. In this case, the disruption may change the direction in which the sleep of one partner shapes and predicts the sleep of the other partner. The analysis of the process by which infants and parents go back to sleep after a night-time awakening, suggested by the models of Chow and colleagues, also is an analysis of how families respond to disruptions. The application of the dual process model presented by Teachman clearly focuses on how developmental change can influence intra-family dynamics. Finally, all three approaches could examine different ways in which family members can influence each other's sleep, a practice traditionally examined in a single individual in isolation from others.

Yet despite these benefits, these three approaches present several challenges. The parallel process model examining concurrent changes in teen and parent sleep would require a fairly typical, but labor-intensive and expensive longitudinal design. The sample size would need to be sufficiently large for enough statistical power, and multiple waves would be necessary in order to have enough variability to estimate the within-family variability across time. The detailed day-to-day variability in sleep patterns necessary for the examination of whether partners act in a self-organizing manner in terms of sleep mandates intensive, daily measurement of sleep. The compliance required for this design would necessitate generous incentives for the participants. The same would be true for the analysis of parent-infant sleep patterns after a disruption during the night, as well as the use of specialized equipment such as actigraphy that can measure sleep on a moment-by-moment basis. The great time and effort necessary for all three designs requires a high level of resources and likely selects for particularly cooperative and agreeable families.

Conclusion

Teachman, Ram et al., and Chow et al. offer unique and potentially powerful analytical techniques to go beyond simply describing the central tendencies of families. The understanding of intra-family variability offered by their models goes beyond traditional designs and at times, even commonly-used analytical approaches to family variability. Yet the data collection efforts required to exploit the full value of their techniques present many challenges. These challenges do not mean that the effort should not be undertaken, but challenges should be balanced with the projected benefits. Doing so would ensure that the value of considering intra-family variability goes beyond family researchers and is effectively communicated to both scientific and lay consumers of the research.

References

Carskadon, M. A. (2002). Factors influencing sleep patterns in adolescents. In M. A. Carskadon (Ed.), *Adolescent sleep patterns: Biological, social, and psychological influences* (pp. 4–26). Cambridge: Cambridge University Press.

Carskadon, M. A., Acebo, C., & Jenni, O. G. (2004). Regulation of adolescent sleep: Implications for behavior. *Annals of the New York Academy of Sciences, 1021,* 276–291.

Dahl, R. E., & Lewin, D. S. (2002). Pathways to adolescent health: Sleep regulation and behavior. *Journal of Adolescent Health, 31*(Suppl 6), 175–184.

El-Sheikh, M., Buckhalt, J. A., Mize, J., & Acebo, C. (2006). Marital conflict and disruption of children's sleep. *Child Development, 77,* 31–43.

Fuligni, A. J., & Hardway, C. (2006). Daily variations in adolescent sleep, activities, and mood. *Journal of Research on Adolescence, 16,* 353–378.

Gillen-O'Neel, C., Huynh, V., & Fuligni, A. J. (2013). To study or to sleep? The academic costs of extra studying and sleep loss. *Child Development, 84*(1), 133–142.

McHale, S. M., Kim, J. Y., Kan, M., & Updegraff, K. A. (2011). Sleep in Mexican-American adolescents: Social ecological and well-being correlates. *Journal of Youth and Adolescence, 40*(6), 666–679.

Meijer, A. M., Habekothé, R. T., & Van Den Wittenboer, G. L. H. (2001). Mental health, parental rules and sleep in pre-adolescents. *Journal of Sleep Research, 10*(4), 297–302.

Part II
Family Systems

Chapter 5
Anatomies of Kinship: Preliminary Network Models for Change and Diversity in the Formal Structure of American Families

Robin Gauthier and James Moody

> *Probably the only way to give an account of the practical coherence of practices and works is to construct generative models which reproduce in their own terms the logic from which that coherence is generated. (Bourdieu 1990, p. 92) Mother is a verb, not a noun.-Proverb*

Introduction

In the face of rapid and fundamental changes, social scientists and the public alike are grappling with how to define "family" and "kinship" in a substantively meaningful way. Families have traditionally been rooted in biological relationships that seamlessly carry *social* roles–the rights and responsibilities associated with our shared understandings of family. But the linkage between such socially defined roles and biological status is not given in settings (such as ours) where biology is only part of the relationships comprising families. In fact, there is no necessary connection, and roles may be completely divorced from biology (White 1963) Since fundamental changes to family structure over the last 40 years have driven a proliferation of new family forms, we need tools that allow us to map kinship as a direct social practice that is not necessarily linked to biological understandings of kin. Here, we explore using an old tool, formal algebraic relational models, in this new context. Our goal is to explore the promise of these approaches and set the foundation for a new style of grounded empirical models of family forms in the future.

Traditional categorical approaches to family definitions predefine important relationships such as marriage or parentage and indicate whether a given household contains such a relationship while ignoring the larger configuration of roles in the household (Widmer and La Farga 2000) This archetypical reduction presupposes "the family" and pushes researchers to focus on "deviations" from it. In this vein,

R. Gauthier (✉) · J. Moody
Department of Sociology, Duke University, Durham, NC, USA
e-mail: grg5@soc.duke.edu

S. M. McHale et al. (eds.), *Emerging Methods in Family Research,*
National Symposium on Family Issues 4, DOI 10.1007/978-3-319-01562-0_5,

scholars of stepfamilies argue that nuclear families are held up as the standard, ignoring alternative family forms and important processual differences among them (Stewart 2007) In response there have been calls for relational approaches that allow for a more organic picture of family roles to emerge (Scanzoni and Marsiglio 1991) Our use of formal kinship tools fills this gap. By building on traditional kinship methods that understand families as systems of overlapping role sets, we can extend these concepts to modern network methods.

In what follows, we first discuss the history of formal structured kinship models emerging from structural anthropology. We then illustrate how these tools can inform our understandings of current trends by discussing the implications of demographic changes for family composition in China and the US, showing that the availability of kin (composition) implies root changes to the role system. While this example is informative for tool building and provides a clear deductive case for explication, it does not substantively move beyond the standard biology-based models. The true promise of the formal role-structure approach emerges in our second example in which we explore the possibilities of an inductive approach to derive roles-as-practiced from time-use data. The application of this approach allows "families" to emerge from the patterns of shared time, indicative of overlapping roles.

Theory: Status, Role and System

In 1936, Linton defined status as "a collection of rights and duties" (Linton 1936, p. 113) and role as "the activation of a status" (Linton 1936, p. 114). Thus, status is a position in an organized system characterized by social expectations, distinct from its occupants, and a role is the active relationship between statuses. An individual *holds* status, and *enacts* a role (Nadel 1957, p. 11), an idea we are all familiar with in organizations. (e.g. The title "assistant manager" is distinct from the individual who holds the position.)

Our tacit knowledge of how a business organization works and how positions relate to one another inform the specifics of any particular organization. The concept "assistant manager" is only meaningful within the framework of this understanding. Our understanding of a formal organization rests on the simple relations of "reports to" and "branch"–a clear rule for how authority flows in a system of offices. These two fundamental relations ("reports to" and "same jurisdiction") define formal organization and are transposable across multiple organizations. The specific duties of a particular office (typing, accounting, billing, etc.) are superficial fillings for the ravioli of the office: the dish is defined by the structure, not the content. Our fundamental argument is that "family" is exactly analogous: family is defined by a complex of role relations, fulfilled by the activities that comprise familial interaction and care.

Families and the kinship systems they are embedded within are also systems of positions systematically related to each other. There is no way to explain what a cousin is without reference to other kinship terms (i.e. aunt, uncle) which themselves refer to others (mother, father, sister, brother, child) rooted ultimately in the concepts "is a parent of" and "is paired with/married to." Yet, intuitively we know this: brothers are brothers regardless of having one or ten; sisters are "like" brothers, while cousins are more similar to siblings than to uncles, but are somehow not the same. The systematic nature of these equivalencies is socially defined by the expectations built into the context for social obligations across roles.

The traditional foundation relations for kinship systems are gender and relative age (generation). Patrilineal descent systems join the child to the father's group, matrilineal systems to the mother's, and bilateral systems acknowledge both paternal and maternal contributions. Political membership and social rights are traditionally passed through different lines. In most societies, kinship *terminology* reflects meaningful social distinctions. In western kinship, a cousin is ego's parent's sibling's child, but this designation is a product of the distinctions underlying our kinship system (gender of parent and child is irrelevant but generation is not) and "cousin" as such need not be acknowledged universally. Western kinship systems do not strongly differentiate between male and female lines, nor does our terminology[1]. We do, however, distinguish between full siblings and step-siblings, suggesting a distinction that is institutionalized and meaningful. Because terminology reflects social distinctions, we can use it to uncover the underlying system logic.

The notation used in this chapter is straightforward (representing concrete social roles-i.e. "mother's brother's daughter" or "father's sister's daughter") and flexible enough to admit non-biological ties. This flexibility also allows researchers to accommodate local knowledge of relationships, such as societies that omit a generational distinction or do not differentiate between father and father's father (White 1963; Bourdieu 1990; Read 2007)

The quantity of relationship types and partitions recognized within a kinship system has a fundamental effect on how complex the kinship system can be. Simpler classificatory systems have relatively fewer roles than systems with several age classes or that distinguish unique roles several steps out. Likewise, the number of people enacting roles from any ego's perspective limits the possible role configurations. If n is the number of people in a household or family tree and k is the number of distinct roles, a household with three members can only have $3!k$ $(= 6k)$ possible role combinations while a household with four members has $4!k$ $(= 24k)$ possibilities.

The beauty of the formal role systems approach is that it allows us to distinguish equivalent elements across settings and thus derive roles independent of standard labels. For example, systems that create a strong division between maternal and paternal contributions typically imply highly differentiated gender roles, yet there are cases when a person of one gender can be fully incorporated into the role associated with the other. The Nuer (people of South Sudan and western Ethiopia) reckoned lineage through the patriline, transferring tribal membership from father to his children. However wealthy women could also have "wives," and high status widows

[1] While we do formally have "maternal grandmother" or "uncle on my father's side" the former is rarely used in everyday family and the latter, used more commonly, highlights the very lack of an everyday term.

were effectively men, overcoming biological constraints to continue the logic of inheritance. Any children born to the wife would belong to the lineage of the female husband (Evans-Pritchard 1940) Robert Brain (1972) documented female fathers among the Bangwa (people of west Cameroon) mainly among high status women. These women could inherit wives. Any children born to these wives would inherit from the female husband just as they would from a male father. Such gender-breaking family roles give us an indication of how to make sense of the seemingly fractured world of the contemporary American family (and is implicit in our understandings of ideas like "Mr. Mom"): Focus on the behavior pattern over the biology.[2]

While some systems assign different inheritances (political and social) from each line, others question the biological nature of parenthood. Malinowski (1932) finds among the Trobrianders of New Guinea that one does not become a father through sexual intercourse, but rather through providing the mother's womb with an idea that will become the child. We can see a similar idea in the history of our own legal system with the distinction between biological fatherhood and legal fatherhood: Unmarried fathers must typically apply for their legal rights and responsibilities to be recognized (Guzzo 2009) Adoption presents another example–adopted children bear no relation to their adopted family, yet they carry the family name and are expected to act and be treated as biological children. Modern and unique situations such as surrogate motherhood and sperm donation have forced us[3] to think hard about what precisely counts as "kin."

Nonetheless, the dominant linguistic family frame for western kinship structures is biology. Thomson (2005) demonstrates how people using assisted reproductive technologies choose donors and surrogates with care to maximize genetic similarity to substantiate their claim to parenthood. The flexibility of the meaning of kinship becomes clear when parents explain the logic behind asking daughters to be their surrogate mothers, or when they ask that donors be from the same ethnic group, or when they ask their friends to provide an egg. In general, the parents place greatest importance on the biological matter they, themselves will be contributing to the child, separating the social connection between "birth" and "parent."

The preceding examples show how deviations from typical role enactment at once challenge and uphold the kinship system. Like a transposition, these variations are "out of tune" but do not fundamentally alter the melody. For kinship to serve as a fundamental organizing feature within communities, members of that community need shared meaning, which implies a systematic logic of kin inclusion. Just as grammar guides everyday speech even when respondents are unable to specify the

---

[2] Western kinship does not readily distinguish maternal from paternal biological or social inheritances, yet maternal rights are automatic while paternal rights require legal intervention via marriage or non-marital custody arrangements.

[3] "us" being researchers: This is a much more vexing problem for researchers than practitioners. It is trivial for children to name the relations of their kin. Researchers see confusion where children see clarity because the researcher is looking for a universal frame to cover all settings; while the child cares only for the consistency of his or her own local system. In a modern context, multiple local contexts might be quite distinct from each other, which is the central break from traditional structural anthropology which has been criticized for its insistence on one dominant frame.

linguistic rules, kinship systems must have an underlying commonality to be sensible. As long as the underlying logics of kinship roles are well defined and stable, it is helpful to use a formal analysis of terminology. If however, underlying logics of kinship have changed, an approach that views terminology as a complex behavioral grammar suggests that deviations from the terminologically prescribed behavior are grammatically incorrect. We agree with Stack and Burton (1993) who argue that there is no such thing as "the family," but rather that families are localized systems that respond to local economic and cultural needs. We agree with Bearman (1997) that a behavioral approach to define roles may be used to uncover these logics from patterns of interaction.

To help think through this logic, we make use of two important distinctions raised by Nadel (1957) that are relevant for the kin term system in the US and China. First he distinguishes structure from content and then goes on to argue, "One type of structuring is abstracted from interactions, the other from distributions," (Nadel 1957, p. 15). While kin term analysis focuses on how kin terms relate to each other, network analysis works backwards from interaction to *derive* kinship. These two distinctions mirror our analyses below. First, we use the formal relationships between different kin *term* systems in China and the US. This makes clear the logic behind kinship systems and how seemingly quantitative changes have fundamentally qualitative effects. We then reverse the process and describe how one could map American household structure based on what we currently know about the US kinship system with an empirical example. Finally, we suggest a new method to derive behavioral roles from interactions by identifying patterns in time spent together.

Formal Roles From Kinship Terms

We draw on a long tradition of formal analysis that uses a mathematical representation of a system to foster comparability. We owe a particular debt to White's (1963) *An Anatomay of Kinship* which outlined a method to reduce family types to their generative (primary) rules and examine the structure that results from the way the primary roles cumulate. The intuition here is had easily by generalizing traditional kinship extension terms, then reducing those to Boolean compositions in a series of family-structure equations. For example, we know that Z = MD; "sister equals mother's daughter" while "cousin" equals my parents-siblings-child. White's insight was that instead of compiling long sentences of such relations (MFZD—mother's father's sister's daughter), we can represent each relation as a matrix of equivalencies and induce such relations by multiplying across the sets of relations. In some cases, this multiplication yields an identity; the spouse of my spouse is me, or the sibling of a sibling is a sibling. These final equivalencies form the boundaries of the system in which strings of compositions yield nothing (empty/full matrix) or only repetitions of relations we know. In so doing, you can generate the entirety of a kinship system with a small number of primary relations, We draw on White's model to identify which relationships are necessary to fully represent the modern American kinship system.

Table 5.1 Examples of common western kinship terms within one and two step neighborhoods

One-step neighborhood	Two-step neighborhood	Kinship term
Father		Father
Mother		Mother
	Mother of mother	Maternal grandmother
	Father of mother	Maternal grandfather
	Son of father	Brother

Practically, this is achieved by identifying relationally equivalent role sets from the data. Role sets are relationally equivalent if they are always found in households with the same role profiles (White et al. 1976) In answering the question, one can derive the underlying logic of how the system is ordered and possibly uncover social rules that even the participants are unaware of.

Applications

We now apply some of these ideas in three cases. The structure of kinship is constrained by the distribution of available kin types and we can use structural analysis to observe the organization of kinship when the system is changing. First, we examine how the Chinese one-child policy provides an example of the complex interplay between structural forces and kin types. We then apply a similar logic to the availability of kin in the US. Finally, we examine the effectiveness of using patterns of time use for identifying emergent kinship structures within households.

Structural Limits on Kin Availability: Comparing the US to China The Chinese one-child policy offers a natural context in which to demonstrate the connection between the availability of kin and the types of family structures that can emerge[4]. Family roles in Western kinship derive from two primary relations: “is married to” and “is a parent of”, both of which are further subdivided by gender.

If each term required to describe a relationship is called a step and we assume for simplicity each married couple has two children, the traditional Western kinship model implies a family system encompassing 26 people within two steps. Table 5.1 illustrates these terminological divisions with common examples. Since the number of roles grows quickly with each new member, the number of people in a two-step neighborhood increases to 46 if each reproductive node has three children.

The Chinese kinship system expands the Western model by adding relative age and kinship order to the foundations of “is married to,” “is parent of,” and gender. This means that relations which are equivalent in the Western kinship system, such as aunt and cousin, are differentiated in the Chinese system, expanding our 26 term family system to seventy-four unique kinship terms.

[4] Here we are focusing on the *logic* of a one-child policy—how it changes the underlying set of formal kin ties. In practice, the policy has never been fully enforced so the results on-the-ground are unlikely to be as clear.

Limiting relations to a single child destroys two key compounding elements. First, we no longer have "older" or "younger" as a relevant partition, radically reducing a fundamental kin component. Second, the lack of siblings means that extension of any lateral kin terms (cousin, uncle, or sibling) and their extensions built by compounding over generations disappears. The effects of the one-child policy on this system are profound, stripping fundamental cultural elements from the community. To the extent that kinship forms a social foundation–a scaffolding upon which to build community and social order–it is hard to understate the radical nature of such a reorganization.

Such changes, of course, are not exclusive to dramatic externally imposed social policies. General demographic transitions, such as the current trend toward below-replacement fertility will have similar effects, but such transitions were historically generated through a more gradual endogenous process that allowed the social system to adjust. While it is common to think about the economic effects of such transitions, the effects on cultural schema are also dramatic: Over time entire swaths of the cultural kin-space are left null—simply unoccupied due to the lack of people.

Family Transitions in the US: Patterns in Complex Kin Well-known trends in the components of families create the constraints that drive varieties of family structures. For example, cohabitation has markedly increased since 1980 (Bumpass and Lu 2000) , as has age at first marriage since 1970, so that *if* people marry they do so at later ages. There has also been a dramatic increase in divorce since 1960 (that has since leveled off). and non-marital childbearing has increased continuously since 1960 (Lesthaeghe and Neidert 2006) These trends create a growing number of stepfamilies and non-residential parents, fewer births, and greater residential instability. Finally, despite an increase in average age at first birth and increased births at advanced ages (Billari et al. 2007), increased longevity makes it more likely that older and younger generations will be alive at the same time (Swartz 2009; Bengson 2001).

Together these trends weaken traditional boundaries around the nuclear family defined as two married parents and their shared children. Children commonly link households after marital transitions and multiple generations are more likely to link households when children split off to form their own families. Each of these changes has been documented independently and there is some understanding of their effects on household structure and the content of relationships within the household (parenting style, sibling rivalry and so forth). Most of the focus has been on the prevalence and diversity of stepfamilies. However, these changes are systemically interdependent; shaping the system of relations we call "kin."

In what follows, we use Read's (2007) reduction of the American system of kinship terminology in two partitions; type (common descent or marriage) and generation, further subdividing these by gender. Then we record the presence or absence of each of these terms and cluster household configurations. This accounting generates several distinctly modern family types alongside a surprising array of family forms typically associated with pre-industrial society.

To bridge between models of kinship that use formal, pre-defined, culturally understood kin terms and inducing modern, perhaps un-recognized kin relations, we need to map families as networks. Kin terms are cultural representations of biological

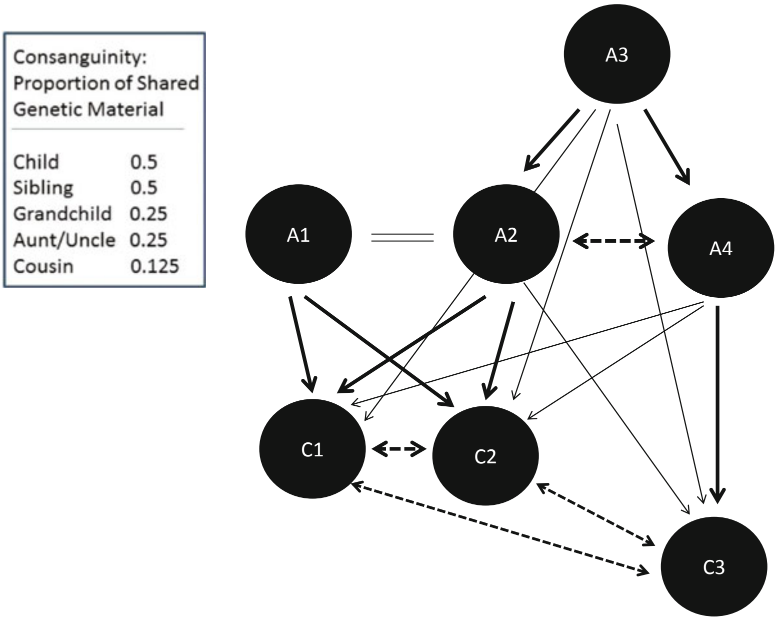


Fig. 5.1 Genetic density in a three generation household

relations which can be mapped as networks. In a P-Graph analysis (White and Jorion 1992) adults are linked to each other and their children through ties of blood or marriage, and the relationships between children are implied by their relationship to the adults. Figure 5.1 provides an example of the resulting network. The two types of relationships that constitute American kinship are represented as arrows and the equal sign. The thickness of the lines connecting the nodes indicates the strength of their biological relationship with the downward facing arrowhead indicating descent, the double headed arrows indicating sibships, and the equal sign indicating marriage.

Figure 5.1 shows three generations with the grandparent at the top [A3], two children [A2 and A4], and A2's spouse [A1]. The married couple have two children [C1 and C2] while the other adult child has one [C3] as well. The light double-headed arrows between the children denote collateral kinship–they are siblings and cousins. We use the weights of the ties to inform our measure of relatedness at a household level–what we call "genetic density". It is a valued network density (Wasserman and Faust 1994) measured as the average value of possible ties. Children take[5] half their genetic makeup from each biological parent. As a result, the most closely related people share about 50 % of their genetics and only two parents can be this close to the children in the household.

$$e = ((N\,adults^{*}N\,adults) - N\,adults)^{*}0.5 + 2^{*}N\,Children^{*}0.5 + (N\,adults - 2)^{*}N\,children^{*}0.25.$$

[5] We conceptualize the genetic tie as asymmetric because children are a product of their parents' genetics.

Table 5.2 Genetic weight matrix (W)

	A3	A1	A2	A4	C1	C2	C3
A1	0	NA	0	0	0.5	0.5	0
A2	0	0	NA	0	0.5	0.5	0.25
A4	0	0	0	NA	0.25	0.25	0.5
A3	NA	0	0.5	0.5	0.25	0.25	0.25

The matrix below is the weight matrix (w) that is a direct transformation of Fig. 5.1 and records the observed genetic material shared between each household member (Table 5.2).

$$o = \sum W.$$

Finally we divide the quantity of observed shared genetics from the possible shared genetics to obtain genetic density at the household level.

$$GD = \frac{o}{e}.$$

In this case, there are four adults and three children which generate 12 dyads among the four adults. If all three adults [A1, A2, and A4] were descended from A3, then the strength of the biological ties between the adults would all be 0.5, and the sum of the weights among the adults would be 6, rather than the observed sum of 1. Only two of these adults are parents to any given child, so the first two adults are assigned a weight of 0.5. The maximum weight of other related adults would be 0.25 for relations like grandparents and aunts/uncles. So all weights between the remaining two adults and the three children are set to 0.25, and the sum of all the weights would be 1.5. Thus, the possible weight of shared genetics in this household is 10.5. However, since not all adults are siblings, nor are they all children of the same parents, the actual sum of genetic weights is 5 and the ratio of shared genetics to possible shared genetics in the example household is 0.47.

Biological relatedness pervades our understanding of kinship and thus the genetic density of households provides us with useful information about the field of possible family structures. In particular, identifying the genetic density allows us to distinguish equivalent roles with different biological foundations, such as comparing cohabiting households with children to married couples. This difference makes it especially interesting to see what other kinds of roles are associated with cohabiters that are not associated with married households or vice-versa. Our idea is to paint a picture of the distribution of US families with respect to both roles and biology (within the limiting factor of household size), as these provide the foundation of any kinship system.

US Context

How are families distributed across kin terms and biological relatedness in the US? The kinship term and genetic density measures defined above provide the tools we need to paint a structural portrait of the contemporary American kin system. On one hand, we're interested in putting together a picture of household configurations such that we know which kin terms tend to correspond with others (and perhaps as importantly, which do not). On the other hand, we are interested in the genetic density of these configurations because it allows us to see how we understand the legal distinctions that go with family formations. Married households with children are indistinguishable biologically from cohabiting ones, yet the two differ socially (Bumpass and Lu 2000). While these two may generally be isomorphic, they need not be, and differences are telling of shifts in the social system.

We start with family data from the General Social Survey (GSS) (Smith et al. 2011), a nationally representative dataset that has been collected for nearly 30 years. The GSS has asked detailed questions about the nature of relationships as part of its core for the last ten years. Questions ask how each household member is related to both the respondent and his or her spouse or partner (distinguishing between the two), which allows us to distinguish whether a minor child is the biological offspring of the respondent only, the respondent's spouse (or partner) only, or a shared child. Most surveys ask only about relations to the main respondent, which forces the researcher to infer relationships between an unrelated child and the householder's partner. Here we pool all 10 years for a total sample size of 16,971 respondents.

The number of possible role combinations in small households is bounded but large. There are eight role combinations for a household size of three if the household head is married and lives with their spouse and one other person holding one of eight non-spousal roles[6]. If the person is not married, there are 45 possible role combinations for a household size of three. Not all possibilities are equally likely, and we can learn about what structures US households by identifying which *configurations* are most likely to occur, a problem easily solved with cluster analysis.

We begin with what we know about kinship in the US and maintain the distinction between types of relationships; legal (married/unmarried), lineal (is a child/parent of), and collateral (is a sibling of). These relationships result in 19 observed kinship classes, which can be further divided by gender, yielding 38 distinct kinship terms. The presence or absence of each of these 38 roles is recorded for each household. Limiting the model to the presence of roles instead of counts, ensures a focus on role configurations (qualitative differences) rather than family size differences (volume).

After constructing the household rosters, we standardized the variables and calculated canonical variables which were then subjected to a K-means clustering model. The researcher specified the number of clusters into which observations are assigned.

[6] In-law relations are possible but reduce to relationship to the head when all role-dyads are enumerated. For example, if a householder lives with a spouse and parent-in-law, the role set contains {spouse, parent-child, parent-in-law-child-in-law} and if a householder lives with a spouse and parent the role set contains {spouse, parent-child, parent-in-law-child-in-law}.

Then initial cluster seeds were chosen by the algorithm as a first guess at these clusters based on the variable means. Observations that were closest to the initial seeds based on Euclidean distance were added to that cluster until there were no observations remaining. A small local peak was observed in the Cubic Clustering Criterion at a 10 cluster solution and the improvement to R squared dropped off quickly as more clusters were added. The family types that were placed into the same cluster are shown below. We restricted our analysis to household types that made up at least 2.5 % of the cluster to avoid enumerating a description of individual households, but we do give an account of their commonalities (Table 5.3).

The first five clusters contain relatively new family forms including living alone or with unrelated roommates, living with children but with no spouse, and cohabiting with and without children. The second set of clusters includes more traditional family forms, including married couples with and without children and a small number of households that include extended kin.

The first cluster is composed mainly of single men and makes up 17 % of households in our sample. The vast majority of these men live alone but just under 10 % of them live with either a son or daughter. Twelve percent live with some other assortment of relatives, mostly their mothers and siblings. The second cluster is composed of single women. It mirrors the first cluster of single males in size with 17 % of the sample being single women. However, there are important gender differences. Single women are less likely to be living with roommates and other relatives. When single women do live with other relatives, they are less likely to have their children among them–none of the women in this cluster are living with their own children. The third cluster is made up of single women with children and grandchildren. This cluster is considerably large, containing over 10 % of our sample. There are no clear gender differences among the children. Between a quarter and a third of households in this cluster contain a mother/daughter pair, a mother/son pair, or a mother/daughter/son triad. The remaining 18 % of the sample are various household forms containing a daughter, her children, and her mother.

The fourth cluster makes up about 4 % of the sample and contains generally isolated cohabiting couples. Ninety percent of the couples live alone and the remaining 10 % live with other relatives, mainly brothers. Cohabiting couples with children make up the fifth cluster. It makes up a considerably smaller portion of the sample (2.74 %) than the married cluster and is more heterogeneous. These households are proportionately much more likely to contain stepchildren.

Ninety-five percent of the sixth cluster is made up of an isolated married couple. The remaining 5 % of households in the sixth cluster consist of a married couple but also other relatives. The most common other relative is a grandchild. Together the sixth cluster makes up almost a quarter of the sample. Households with a married couple and children are equally prominent. The seventh cluster contains married couples and their children. Both stepchildren and shared children are included in this cluster. Slightly more households in the seventh cluster contain only sons (29 %) than daughters (24 %), and more contain both (33 %). The remaining 13 % contain a mix of relatives including stepchildren. None of these unique configurations characterize more than 2.5 % of the households in the seventh cluster.

Table 5.3 US household composition

Single men	Single women	Single mothers	Cohabiting couples with no children	Cohabiting couples with (Step) children
73.5 % male alone 3.3 % male with daughter only 4.8 % male with son only 6.4 % male with roommate only Remaining 12 % is a combination of assorted relatives, children	88 % female alone 3 % female head, female roommate remaining 9 % assorted relatives, female head	28.5 % female head, daughter 30 % female head, son 24 % female head, daughter and son The remaining 18 % are permutations of female heads, their children and grandchildren	90 % cohabiting couples living alone The remaining 10 % assorted roommates and relatives, especially brothers	16 % cohabiting couple and shared daughter 3 % cohabiting couple and shared daughter and male other relative 20 % cohabiting couple and shared son 12 % cohabiting couple, son and stepson 13 % cohabiting couple, shared son and daughter remaining 36 % are other relatives, especially stepchildren
Married couples with no children	Married couples with (Step) children	Collateral families with no children	Collateral families with children	Stem families
95 % married couples 5 % assorted family, esp. grandchildren	25 % married couple and daughter 29 % married couple and son 33 % married couple with daughter and son 13 % assorted relatives including stepchildren	30 % married couple, brother/brother-in-law 10 % married couple, brother/brother-in-law sister/sister-in-law 10 % married couple, daughter, brother/brother-in-law 15 % married couple, son, brother/brother-in-law 10 % married couple, son, daughter, brother	All 31 households are different permutations of married couples and their collateral kin/own children and nephews	31 % married couple and mother/mother-in-law 4 % married couple and mother/mother-in-law, father/father-in-law 17 % married couple, daughter and mother/mother-in-law 10 % married couple, son and mother/mother-in-law 13 % married couple, son and daughter and mother/mother-in-law The remaining 25 % adds grandchildren, male parents/parents-in-law

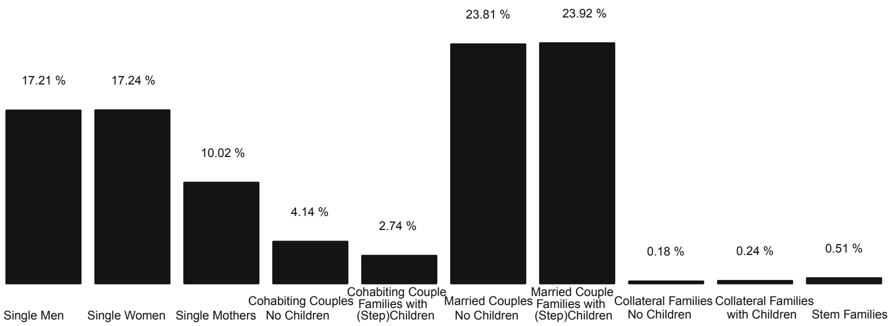


Fig. 5.2 Distribution of households into ten clusters

The eighth cluster is also composed of households headed by a married couple and include their siblings. This cluster contains less than 1 % of the sample. Rather than sisters/sisters-in-law, this cluster contains brothers/brothers-in-law and very few (2) of their children. The households in the ninth cluster are broadly characterized by a married couple and other collateral kin. Most also contain a son and/or daughter and about half include a sister/sister-in-law and her male and female children. Finally, there is a small proportion of families (about half a percentage) that contain a married couple and one set of their parents. These are traditionally called stem families and are more common in preindustrial settings. The distribution of household across these ten clusters is shown below (Fig. 5.2).

This analysis may appear to contradict early studies that show far higher proportion of cohabiters, but it is important to note that this data is a cross section of the entire US and does not over represent younger people forming families.

The genetic density of each household was calculated by transforming the membership roster into a network of genetic relationships weighted by the proportion of shared genetics, implied by the kinship term used to represent the relationship. As in the earlier example, relationships are directed (parents to children) and relations between children are implied. Finally, the weight of realized relationships was divided by the possible weight of relationships. As plot 3 below indicates, genetic density ranges from 0 to 100 % among US households with the many households sharing no genetics. These households may be couples with no children or roommates (Fig. 5.3).

Figure 5.4 below provides the genetic density within each cluster. Households clustered together because they contained no children (the cohabiting cluster and the roommate cluster) have the lowest genetic density. Household containing no children, but containing siblings and their parents (the collateral cluster) have the highest genetic density. The range of genetic density is highest among the cohabiting with children group. This group contains both shared children and children from previous unions. Genetic density is also quite variable among the collateral kinship groups because at least some relationships imply a high level of relatedness due to sharing a common ancestor, while others are legally defined.

Only the presence of roles was used to identify household clusters. The presence of one parent-child dyad was treated the same as three parent-child dyads because we

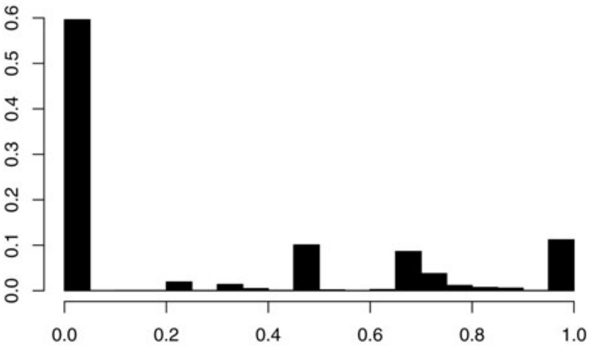


Fig. 5.3 Distribution of genetic density across US households

wanted to cluster unique role configurations, and distinguishing the number of each kind of dyad offers no new information on the role composition of the household. However, genetic density was calculated based on all dyads in the households. As a result, the number of dyads, and thus the size of the household, shapes the level of possible genetic density in each household.

To summarize our findings, differences based on gender and on the legal status of adults are apparent. Single women are more likely to live with children than single men. While single women who live with their children are classified into a distinct cluster, single men who live with their children are not. These men, in fact, have household membership profiles that are more similar to the patterns found among single men with no present children. Single men, both those living with and without children, tend to live with roommates and other relatives, much like single women. However, when single women live with their children they are less likely to live with other family members (except their own mothers) than are single men who live with their children. Households containing cohabiting couples have more diverse role configurations than households headed by a married couple, whether or not there are children present. This suggests that there is no profile for cohabiters–they are in the middle of the size distribution and the boundaries around their households are more permeable than married households.

Families as Overlapping Role Sets

The ultimate promise of a formal network approach is to free us from relying on labels entirely, allowing us instead to use behavioral information to induce roles directly. In this last section we introduce what that process might look like. We want to use networks of shared activities to characterize role sets from an individual perspective. An individual's role set is the dual system of the set of actions they participate in throughout their day and the people who participate in these actions with them. This is a particularly appealing approach when social systems are unsettled or in a state of flux, as individuals may be unable to clearly articulate the behavioral expectations

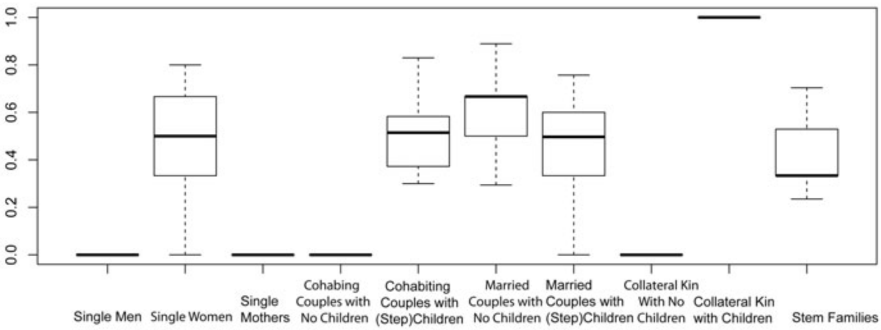


Fig. 5.4 Genetic density by household configuration type

(or implications) of a label. Even when society is clear about which labels apply to which people, it is possible that privileging respondents' local understanding of the system to derive the rules guiding the macro structure can be misleading (Bearman 1997) For kinship to serve as a fundamental organizing feature within communities, members of that community need shared meaning, which implies a systematic logic of kin inclusion. Just as grammar guides everyday speech even when respondents are unable to specify the linguistic rules, kinship systems must have an underlying commonality to be sensible.

In what follows, we provide a look at some basic structural features of the enacted role sets that constitute American families. Specifically, we measure the complexity and differentiation of family role sets. Role complexity is defined by the number of different kinds of actions. For example, a mother is likely to have a complex role set if she does a wide variety of things with her children and even more, different things with her friends. Role differentiation is defined by the number of different people the activity is shared with. A person who has many relations which are narrowly defined with a single focus (professional, socializing/relaxing) will have a differentiated role set.

To map the behavioral portrait of families, we employ time-use diaries. Time-use diaries record detailed data on everyday activities, including who people are with and whether they were active participants or merely present (Paolisso and Hames 2010). Time-use diaries allow us to construct a network of people through activities, reflecting the realized practice of family life as who-does-what-with-whom. We can then use bipartite networks of relations, not through names, but through practice, to categorize families' structure directly by shared activity.

The American Time Use Survey is a nationally representative study of the amount of time people spend doing various activities that has been carried out annually since 2003. It draws its sampling frame from the Current Population Survey (CPS), a monthly survey of households conducted by the Bureau of Census. Respondents are randomly chosen from the households but must be over 15 years old. No substitute proxies were allowed. The sample was divided into four panels, each assigned to one of either weekend day and the other two divided among the week days to avoid "work-week" variation. Additional questions are asked about "overnight" trips since those tend to be missed. This sampling frame is particularly advantageous for our purpose because it contains weekends and overnight trips and is therefore likely to catch visits between non-resident parents and their children. For most activities reported, respondents are asked "Who was in the room with you" or "Who accompanied you?" Each household member and non-household child is assigned a separate "who" code. Generic categories also exist for non-household family members and for others (e.g., neighbors, friends).

The time use data provides a rich record of respondents' daily experiences. As an example, consider stylized response set derived heavily from one respondent's diary. The respondent is an adult female who described spending a Sunday with two other people-her child and her cohabiting partner. She reported having a rather typical day. She woke up and got dressed then did some grocery shopping with her partner, then came home and made lunch. She ate lunch and cleaned up by herself and then went on to clean up the house and do the laundry with her partner. Then she spent time talking with her child. That evening she relaxed by smoking and eating dinner with her partner and watching TV with her partner and child. Finally, she and her partner went to sleep. We use the seven aggregated categories provided by ATUS to simplify the following analysis: (1) household activities, including cleaning and meal preparation, (2) travel, (3) shopping, (4) care for another household member, (5) personal care, (6) socializing and relaxing and (7) eating.

The data can be read as an itinerary and this list of activities can tell that the set of things the respondent does with her partner is different from the set of things she does with her child. We refer to the respondent's partner and child as her alters. They are the set of people accompanying her though the day. While she got ready for the day, ate and smoked with her partner, she helped her child. On the other hand, some activities were shared: She went to the grocery store and relaxed with both her partner and child, and both partner and child were recruited to assist with housework at some point in the day. The role of partner is manifestly different from the role of child, and moreover in the life of this respondent the partner role is more complex because together they did more different things than she did with her child. The family did many things together and so her role set is not particularly differentiated despite the wider range of things done with her partner. We focus on these two rather simple characteristics of the role sets embedded in their networks that summarize important features of their day-to-day lives.

We represent the respondents' alters and their shared behaviors as a bipartite network to facilitate further analysis. A bipartite network represents the duality of persons and groups (Breiger 1974; Feld 1981) by mapping implicit social ties. A network is bipartite if there are two sets of nodes (alters and behaviors) where all ties are only found between sets (Wasserman and Faust 1994). The main advantage of adopting this framework is that statistics developed for use on traditional one-mode networks can be modified and applied to two-mode networks. Figure 5.5 is an image of the selected respondent's day. The respondent's alters are represented along the

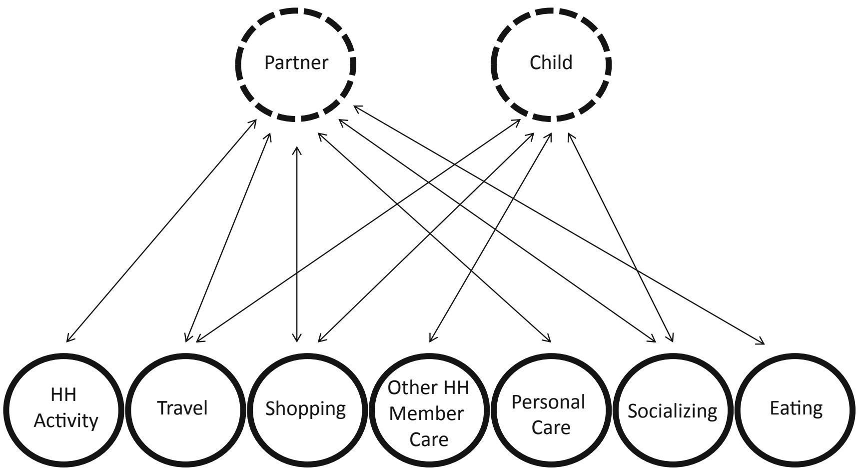


Fig. 5.5 Bipartite representation of respondent's day

top and the activities she does with them are listed in the bottom. Alters are connected to activities but not directly to each other.

We use this representation of time use networks to calculate statistics for all the respondents in the next section. We operationalize the complexity of each respondent's role set by taking the average number of unique activities they report doing with each alter. This measures the average complexity of the respondent's role set. We use the Jaccard dissimilarity index (Oksanen et al. 2012) of the respondent's alters to measure role differentiation. The Jaccard index is a measure of dissimilarity between a pair of vectors. It ranges from 0 to 1 and is higher when the respondent's alters have different activity sets and lower when they do more of the same things with the respondent.

We begin by analyzing role set complexity by age in the first panel of Fig. 5.6 below. On average, respondents report 3.5 distinct actions with each alter. While the differences are minor, middle-aged people have the most varied interactions, followed by the young and the elderly respectively. As shown in the second panel, women have slightly more complex role sets than men. Finally, the presence of children increases role complexity by half an action. This is a significant change because the unique actions are averaged over all alters[7].

Next we turn to role differentiation in Fig. 5.7. The first panel indicates that role sets have more overlap among young and middle aged respondents. Respondents over 65 years old have more differentiated roles, suggesting they are more likely

[7] Gender and child differences are statistically significant at conventional levels and remain significant when all variables are added to the model. The age contrast between 65 + and middle-age are significant. In addition to the variables on the figure, we control for the number of social actions the respondent reported overall and the number of people the respondent reported interacting with overall.

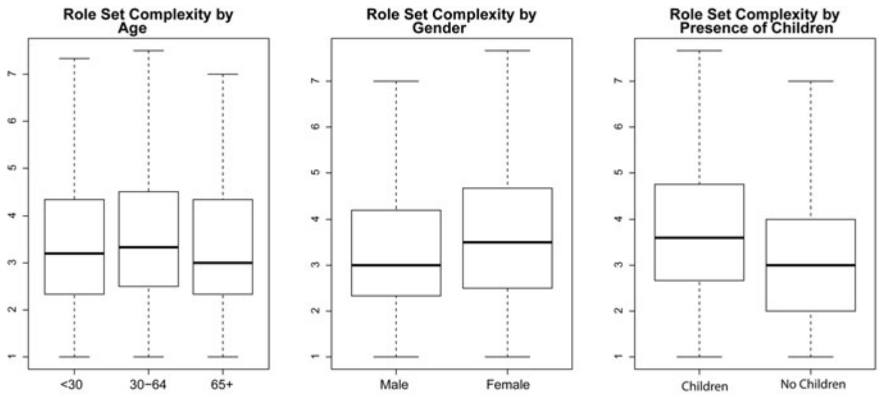


Fig. 5.6 Role complexity by respondent characteristics

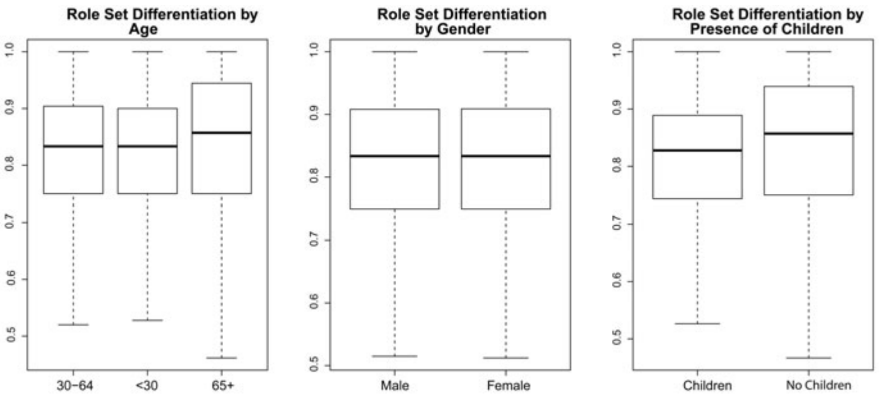


Fig. 5.7 Role differentiation by respondent characteristics

to specialize in who they do things with. The second panel shows that there is not much difference in role differentiation by gender. Finally, children decrease differentiation[8].

Gender differences in both complexity and differentiation are small. However, the roles in women's daily lives on average entail more activity than the roles in men's daily lives. Nonetheless, the roles women enact are as similar to one another as men's are. Children's needs structure their caregivers' actions, requiring a complex set of caring actions. On the other hand, the presence of children raises expectations of spending quality time together which would decrease the differentiation of roles in the household.

[8] Age and children differences are statistically significant at conventional levels. However, age drops from significance when the presence of children is added to the model.

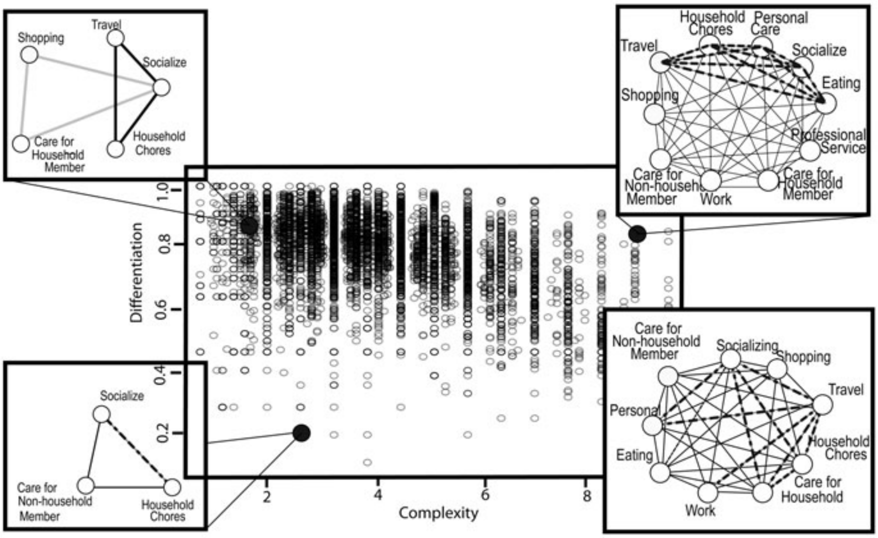


Fig. 5.8 The relationship between role complexity and differentiation

To get a sense of the range of role sets, we array them by complexity and differentiation in Fig. 5.8 below. This space defines a field of possible experienced roles. A main diagonal flows along the general negative correlation, anchored at one end by highly complex but low differentiation networks (lower-right) and highly differentiated but low complexity networks in the upper-left.

Off this diagonal, low complexity is coupled with low differentiation and high complexity is coupled with high differentiation. We argued above that kin-type relationships are relatively highly structured. Relationships outside the bounds of kinship (e.g. friends or coworkers) may be less complex and differentiated if they do not include the complex demands of kinship. If, for example, most of the social things a person does are within one highly circumscribed social setting like work, or if the relationships they have are less structured and focused around a single activity like casual friendships and acquaintances, the relationships would entail more unidimensional activities. We expect then, that while not all non-kinship relationships will be simple or undifferentiated, people with role sets with low differentiation and complexity are less likely to be kin-based.

Respondents who named more non-kin alters in their daily interactions are more likely to have undifferentiated and non-complex role sets, while respondents who interact with children are more likely to have role sets that are both more complicated and more differentiated than others.

Conclusion

Current research uses kinship labels as a proxy for relationships, but the growing decoupling of kinship terminology from relationships, and the emergence of new kinds of relations that lack terms, suggest a revision of that practice is in order. Here we (re)introduced an old framework: By extending the logic of traditional kinship studies to contemporary social network techniques we suggest a formal, inductive approach to uncover patterns of family membership. This approach allows us to truly leave behind comparisons to the nuclear family. If the social networks that make up alternative family forms overlap in unique ways, we can see the multiplicity of logics that makes up families and then understand them on their own terms, situated in particular economic and cultural settings.

Early anthropological methods to uncover the kinship systems of foreign cultures using systems of kinship terminology have fallen out of favor because they are seen as static and structurally deterministic, often implying that alternative behaviors are dysfunctional or deviant. We argue that the fundamental logic of these methods is valuable to understanding families as they are today, and if taken as representative of the constraints of localized social systems, rather than normative universals, offer a potentially powerful tool for mapping family diversity across the US. Without assuming a *single* logic, we suggest that family systems develop out of particular social, cultural, and economic environments each operating according to their *own* logic.

Since traditional models of formal ties in families rest on the very elements undergoing change in our current age, we need a way to map roles that is independent of the family names we often rely on. Here, we turn to social practice: Roles are characterized by repeated interactions, so we can induce such relations by looking at how people spend time with each other. This novel approach to family networks through restructuring time use data extends early anthropological methods *to families as they are experienced*. It provides an inductive approach that builds relationship categories from interaction, rather than assume a relationship's presence through the name. The theoretical payoff is the development of a general explanation of kinship variation as social practice that is not necessarily linked to shared genealogy, allowing the researcher to compare respondents' family networks across multiple relation types. Hence the modern family context is characterized by myriad solutions to local structural problems, generating a wide array of kinship forms. In contrast to mid 20th century anthropology models, we live in a world characterized by multiple family forms, and thus have a need for multiple kinship anatomies. While the work here is a first-step, we think the approach presented here holds deep promise for understanding complex distributed families.

Author's Note Thanks to members of the social structure working group, especially Jeffrey Smith, and participants at the Symposium on Family Issues for comments, and to the Boone Fellowship to Gauthier for time supporting this work.

References

Bearman, P. (1997). Generalized exchange. *American Journal of Sociology, 102*(5), 1383–1415.
Bengson, V. L. (2001). Beyond the nuclear family: The increasing importance of multigenerational bonds. *Journal of Marriage and Family, 63*(1), 1–16.
Billari, F., Kohler, H. P., Andersson, G., & Lundstrom, H. (2007). Approaching the limit: Long-term trends in late and very late fertility. *Population and Development Review, 33*(1), 149–70.
Bourdieu, P. (1990). *The logic of practice*. Stanford: Stanford University Press.
Breiger, R. L. (1974). The duality of persons and groups. *Social Forces, 53*(2), 181–190.
Brain, R. (1972). *Bangwa kinship and marriage*. New York: Cambridge University Press.
Bumpass, L., & Lu, H. H. (2000). Trends in cohabitation and implications for children's family contexts in the United States. *Population Studies, 54*(1), 29–41.
Evans-Pritchard, E. E. (1940). *The Nuer: A description of the modes of livelihood and political institutions of a Nilotic people*. New York: Oxford University Press.
Feld, S. (1981). The focused organization of social ties. *American Journal of Sociology, 86*(5), 1015–1035.
Guzzo, K. B. (2009). Paternity establishment for men's nonmarital births. *Population Research and Policy Review, 28*(6), 853–872.
Lesthaeghe, R. J., & Neidert, L. (2006). The second demographic transition in the United States: Exception or textbook example? *Population and Development Review, 32*(4), 669–698.
Levi-Strauss, C. (1969). *Elementary structures of kinship*. Boston: Beacon Press.
Linton, R. (1936). *The study of man*. New York: Appleton-Century-Croft.
Malinowski, B. (1932). *Argonauts of the Western Pacific*. Prospect Heights, IL: Waveland Press.
Nadel, S. F. (1957). *The theory of social structure*. London: Cohen & West.
Oksanen, J., Blanchet, G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., & Wagner, H. (2012). vegan: Community Ecology Package. R package version 2.0-5. http://CRAN.R-project.org/package=vegan.
Paolisso, M., & Hames, R. (2010). Time diary versus instantaneous sampling: A comparison of two behavioral research methods. *Field Methods, 22*(4), 1–13.
Read, D. (2007). Kinship theory: A paradigm shift. *Ethnology, 46*(4), 329–364.
Scanzoni, J., & Marsiglio, J. (1991). Wider families as primary relationships. *Marriage & Family Review, 17*(1–2), 117–134.
Smith, T. W., Marsden, P. V., & Hout, M. (2011). *General Social Survey, 1972–2010*. Storrs, CT: Roper Center for Public Opinion Research, University of Connecticut/Ann Arbor MI: Inter-university Consortium for Political and Social Research.
Stack, C. B., & Burton, L. M. (1993). Kinscripts. *Journal of Comparative Family Studies, 24*(2), 157–170.
Stewart, S. D. (2007). *Brave new stepfamilies: Diverse paths towards stepfamily living*. Thousand Oaks: Sage Publications.
Swartz, T. T. (2009). Intergenerational family relations in adulthood: Patterns, variations, and implications in the contemporary United States. *Annual Review of Sociology, 35,* 191–212.
Thomson, C. (2005). *Making parents: The ontological choreography of reproductive technologies*. Cambridge, MA: MIT Press.
Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. New York: Cambridge University Press.
White, H. (1963). *An anatomy of kinship: Mathematical models for structures of cumulated roles*. Englewood Cliffs: Prentice-Hall.
White, H. C., Boorman, S. A., & Breiger, R. L. (1976). Social structure from multiple networks I. *American Journal of Sociology, 81,* 730–780.
White, D. R., & Jorion, P. (1992). Representing and computing kinship: A new approach. *Current Anthropology, 33*(4), 454–463.
Widmer, E. D., & La Farga, L.-A. (2000). Family networks: A sociometric method to study relationships in families. *Field Methods, 12*(2), 108–128.

Chapter 6
Emerging Methods for Studying Families as Systems

E. Mark Cummings, Kathleen N. Bergman and Kelly A. Kuznicki

For over two decades theoretical and conceptual models that advance organismic or systems metaphors for understanding families have been of interest, including approaches inspired by family systems theory (Minuchin 1985) and the research generated by or consistent with these views (Cox and Paley 1997). This movement reflects recognition of the limitations of models of family influence processes that constrain study to bivariate relations between parents and children, assume unidirectional pathways, and underemphasize dynamic processes. In particular, an emerging concern is with understanding transactional processes that reflect processes continuously moving in both directions over time, and which may include complex patterns of influence in mother-child, father-child, interparental, and sibling relationships over time. Models for transactional processes are valuable for considering multiple influences on development and adaptation and have implications for the design of effective interventions.

For example, Schermerhorn and Cummings (2008) recently proposed a theoretical model of "transactional family dynamics" as an approach for understanding family influence processes as resulting from the transactional regulatory processes of dynamic systems. The aim was to further articulate the multiple ways in which family members influence one another, that is, mutual influence processes occurring within families. Moving beyond limited notions of family process pathways, the aim was to encompass many pathways, and to integrate emerging empirical work on these directions (e.g., Schermerhorn et al. 2008; Schermerhorn et al. 2007). The concept of transactional family dynamics provides a framework for the influence of individual family members on multiple family relationships, influences of family relationships

E. M. Cummings (✉) · K. N. Bergman · K. A. Kuznicki
Department of Psychology, Notre Dame University,
Notre Dame, IN, USA
e-mail: edward.m.cummings.10@nd.edu

K. N. Bergman
e-mail: kbergman@nd.edu

K. A. Kuznicki
e-mail: kellykuznicki@gmail.com

S. M. McHale et al. (eds.), *Emerging Methods in Family Research*,
National Symposium on Family Issues 4, DOI 10.1007/978-3-319-01562-0_6,

on one another, and family-wide influences, including reverse directions of effects (Whiteman et al. 2010).

Utilization of broader, process-oriented approaches to research in the family is imperative to increased understanding of influence processes related to families and children. For example, the family system approach has focused on identifying relationship structures, communication patterns, power distributions, and interpersonal boundaries in the midst of the complex interplay among relationships and individuals within the family as a whole (Cox and Paley 1997; Davies and Cicchetti 2004; Minuchin 1985). Another aspect of organization in families that has been postulated is hierarchical structure. Families are composed of subsystems that are organized in a hierarchical fashion and influence each other (Fiese et al. 2000). In order to accurately capture these complex dynamics, investigators are increasingly concerned with examining family functioning longitudinally (Davies and Cicchetti 2004; Rothbaum et al. 2002), reflecting a shift from cross-sectional research to longitudinal methods (Bradbury and Karney 1993).

Gaps in the Study of Families as Systems: Methodology and Measurement

Although attention is being given to conceptual issues surrounding families as systems, and increasing application of longitudinal research designs, assessment of family functioning at a level that reflects the interplay of multiple family systems remains a challenge. For example, investigators have traditionally focused on mothers' responses in an effort to understand the family; however, there is now widespread recognition that the inclusion of fathers provides a more comprehensive perspective (Lamb 2010). There is considerable recognition of the importance of multiple reporters on family systems, with researchers increasingly including mothers, fathers, and siblings.

However, less attention is being paid to the development of new methodologies for assessing families as systems, especially the development of methodologies that go beyond paper-and-pencil questionnaire approaches. The advantages of using multiple methodologies for assessment are well-established; each method has strengths and weaknesses, with the use of multiple methods holding potential to overcome weaknesses of any individual methodology (Cummings et al. 2000). In addition, observational or other rigorous approaches may be essential to capturing, or even recognizing, the operation of some key family processes. For example, the impact of nonverbal interparental anger was little documented until analogue methodologies were applied to the study of the impact of nonverbal and verbal anger by parents on children. To an important degree, the appropriateness and rigor of methodology may influence the message concerning the impact of family processes on children (Cummings and Davies 1994; see Appendix D, Cummings and Davies 2010).

In this context, there is a need for continual innovation and creativity to advance new methodologies that provide new perspectives and more informative approaches to understanding the impact of family on children. For example, young children may be better able to reflect their appraisals of family relationships in the context of engaging interpersonal games or tasks about family relationships, as opposed to the context of completing paper-and-pencil questionnaires. As another example, observational methodologies may provide more valuable perspectives and greater rigor of assessment concerning the interrelations among multiple family members than is possible based on analogue or questionnaire approaches. However, to some extent, understanding of family processes is limited by the availability of methodologies appropriate to the adequate examination of these processes. Many questions and issues are best addressed by means of multiple methodological directions. Relatedly, there is a continual need for creativity and innovation in the study of family systems. This chapter is concerned with the process of developing and advancing innovation in methodologies for the study of family systems and processes. The goal of this chapter is to present and evaluate several new methods to address gaps in the study of families as systems, including a revised MacArthur Story Stem Battery, new directions in the use of daily diaries, and a new observational coding system for coding triadic interactions.

Methods for Studying Families as Systems: A Revised MacArthur Story Stem Battery

The MacArthur Story Stem Battery (MSSB) is pertinent to advancing study of young children's representational models concerning family relationships, which may contribute importantly to continuity of family influences over time (Cummings et al. 2000). One goal of this chapter is to advocate for the use of this methodology in the study of family systems of influence based on children's representational models. The MSSB is a narrative storytelling technique, typically focusing on children's internal representations of the interparental and parent-child relationships (Bretherton et al. 1990). Notably, revisions of the MSSB (MSSB-R; Cummings et al. 2001) in recent years have expanded the utility of this approach so that it is even more valuable for family systems research. These revisions foster the applicability to the study of families in terms of reciprocally influential subsystems, providing more extensive measurement of children's representations of multiple family systems. For example, the MSSB-R includes stories devoted to assessing threats to child's appraisals of the security of the mother-father, mother-child, and father-child relationships (e.g., Schermerhorn et al. 2008).

The methodology employed in the MSSB merits consideration. In this task, children as young as 4 years are presented with a variety of age-appropriate story stems by an experimenter. Dolls are used to represent each member of the family and are matched to the ethnicity and gender of each person. Similarly, props are employed to help children envision the scenarios, and experimenters use lively voices for each character to bring the story stems to life. Once the experimenter has presented the story, the child uses the dolls and props to complete the story as they envision events would unfold within their own family. These rich narratives are then coded for marital conflict, children's security, parent-child attachment, parenting style, and a myriad of other representations of family relationships.

For example, in one MSSB-R story stem, the experimenter enacts a scenario in which the child *action figure* in the narrative for the story stem asks the Dad *action figure* if he can watch TV. The Dad responds that the child must finish dinner first. The Dad is then shown leaving the room and the child proceeds to leave the dinner table to watch TV. Next, the Dad is shown reentering the room. The experimenter then turns to the *participating* child and asks him to complete the story with what he thinks would happen in his family. From this story, coders can assess caregiver competence, parent-child relationship quality, parental warmth and control, relationship spillover (if child brings Mom into story), and the child's overall felt security. Once coded, the MSSB provides a unique look into family dynamics that may advance the study of families in terms of reciprocally influential subsystems.

Recent findings demonstrate that children's representations of particular family subsystems have implications for their representations of other family subsystems (Cummings et al. 2008). This work supports the finding that relationships which are viewed as highly distinct in terms of children's representations may also be interdependent in terms of influences on each other and children's development (Schermerhorn et al. 2008). Moreover, understanding these mutually influential subsystems may have implications for child development and difficulties sometimes attributed to single family relationships (Schermerhorn et al. 2008).

With regard to conceptualizations of how representations relate to patterns of development, attachment theory (e.g., Bowlby 1973; Bretherton 1985; Bretherton and Munholland 2008) advances that children use models or representations of family relationships as guides in new settings or to help them negotiate social experiences. For example, with regard to risk for psychopathology, insecure representations of the family as hostile, unreliable, and potentially threatening may foster stress. The stress can lead to worries about family stability and personal well-being, as well as motivate maladaptive behavioral responses for dealing with stress that may generalize to multiple settings (Cicchetti et al. 1990; Cummings et al. 2008; Davies et al. 2006; Davies et al. 2008).

Studying the influence of the children's representations of the security of the interparental subsystem is pertinent to understanding families from a systems perspective. Child insecurity in the interparental relationship is hypothesized to develop from child experiences within the interparental subsystem (Davies et al. 2006). Examining the child's perspective on the interparental relationship, and how this evaluation contributes to their own sense of security, provides considerable insight into the family. The mutual influence of representations within the family, including the context of marital conflict, is indicated by research to date in this area (e.g., Cummings et al. 2008; Schermerhorn et al. 2008).

With regard to the mutuality of influences associated with increased destructive marital conflict within the interparental subsystem, children tend to become more reactive and in turn, intervene more in the parents' conflicts. Interestingly, families with children who intervene in the interparental subsystem may experience a decrease in future marital conflict, controlling for initial levels of marital conflict (Schermerhorn et al. 2005; Schermerhorn et al. 2007, 2008). Thus, the interparental subsystem influences children's security and children, in turn, influence parents. Taken together, these findings advance the perspective that children are not simply passive recipients of the negative impact of exposure to marital conflict but rather are active agents in the context of marital conflict, responding dynamically and to some modest degree effectively, over time to the threat of marital conflict (Schermerhorn et al. 2005).

Beyond indicating the influence of the interparental subsystem on children's representations of security, the MSSB also advances study on children's security in the parent-child subsystem. Keeping in mind the dynamic interplay between various family systems, previous research documents that insecurity in the interparental relationship has implications for children's security in the parent-child relationship (Du Rocher Schudlich et al. 2004; Schermerhorn et al. 2008; Shamir et al. 2001). Marital conflict leads to ineffectual parenting (Owen and Cox 1997), which strains the parent-child relationship and is associated with decreases in children's representations of security (Cummings and Davies 2010; Cummings et al. 2000; Shamir et al. 2001). When children witness parents exhibiting hostility toward each other, children may also question whether parents will behave the same way toward them and thus, insecurity in the parent-child relationship may arise (Cummings and Davies 2010; Cummings et al. 2000). Furthermore, considering the bidirectional nature of parent-child relationships (Schermerhorn et al. 2005), the interparental relationship may subsequently be influenced by the child's insecurity (Schermerhorn et al. 2008). Because children's insecure representations of the parent-child relationship can have deleterious effects on children's abilities to successfully negotiate developmental tasks (Sturge-Apple et al. 2008), harmony in multiple family relationships is vital.

The MSSB examines family relationships from the perspective of the child. Viewing the family system only from the parents' perspectives has left a gap in process-oriented family systems research (Schermerhorn et al. 2005). Therefore, through the use of these stories which examine multiple family subsystems from the child's perspective, progress can be made to close these gaps, and developmental processes associated with family functioning can be examined more broadly. Furthermore, this methodology can provide valuable perspectives into two underexplored areas in family systems research: the child's relationship with the father and the child's experiences within the interparental relationship. The MSSB provides a child's view of the mother-father relationship that may be unobservable through other means. Finally, use of the MSSB in longitudinal designs allows for examination of mutual patterns of influence within the family, which cross-sectional designs simply cannot provide. Based on multiple recent innovations towards expanding the MSSB to explore multiple family relationships, a revised MSSB is emerging as a valuable method for family systems research.

Beyond these areas of innovation associated with the revised MSSB, multiple future directions can be identified towards advancing the study of the reciprocal influence of family subsystems. Our research team has been involved in the development of a parenting intervention for divorcing or separating families undergoing custody disputes. For the purposes of this new project, the MSSB-R has been further adapted and piloted to be applicable to documenting responses in non-intact families with high interparental conflict. In this context, during the presentation of story stems, children are shown separate homes for each parent (see Bretherton and Munholland 2008) and marital conflict stories adapted to reflect post-divorce family situations. For example, the original MSSB-R marital conflict story stem in which Mom and Dad argue about the messy kitchen and whose turn it is to clean it up is no longer appropriate in a family that has undergone a divorce or separation. Thus, for this new study, Mom and Dad are arguing about how Mom thinks the kitchen at Dad's house is too messy, while Dad says Mom has no say in the matter. Beyond piloting, we anecdotally have found these revisions to be appropriate, and also strikingly accurate, as multiple child participants have remarked how similar events have happened in their families. Using this technique, children have been shown to express their internal representations of various family subsystems and divorce-related themes. Additionally, in this context, multiple children from the same family sometimes participate. Thus, multiple child perspectives are elicited that may provide information about not only children's internal representations within the interparental and parent-child subsystems, but also insight into sibling differences.

Another direction we have been exploring toward improving assessment of children's representations of family relationships is extending the MSSB-R into adolescence, which we refer to as the Story Stem Battery for Adolescents (SSBA). Although dolls are no longer developmentally appropriate, we have found that adolescents are willing to complete story stems in the context of an interview format. These new story stems for children in the adolescent period were developed based upon the original stems from the MSSB-R to (1) allow for developmental trajectories to be examined across age groups, and (2) reflect stage-salient issues of adolescence. For example, in the MSSB-R, one story stem reflects the child being injured by touching a hot pan of cookies before they had cooled, even after the Mom told the child not to touch them. In the parallel form of the scenario in the SSBA, the child gets injured while having friends over when this has been prohibited. When the child tries to get her friends out before the Mom returns, the child trips and is injured. In both MSSB-R and SSBA stories, the child has disobeyed the parent and has been injured. Upon Mom re-entering in the story, we can see whether Mom opts to address the disobedience (with a punishment) or the injury (by tending to the wound) first. We also learn a great deal about Mom's competence, parental warmth and control, and the child's security in the mother-child relationship. Using this new, innovative method along with the MSSB-R, children's internal representations of their familial relationships can be ascertained across a wide age span, ages 4 through 17.

Methods for Studying Families as Systems: Advances in the Use of Daily Diaries

Daily diaries provide a valuable approach for the study of family processes as they occur in a natural setting, such as the home. Another goal of this chapter is to advocate for the use of this methodology in the study of family systems of influence in the home. Diaries can be collected through a variety of mediums, including traditional paper and pen methods and more advanced technological strategies such as electronic devices with which participants can be prompted to respond to a series of questions or record their experiences. Different research designs require different kinds of diary methodologies, including event-contingent, interval-contingent, and signal-contingent methods. For event-contingent methods, participants are asked to record a diary entry as soon as a specific kind of event occurs (e.g., at the completion of each family conflict episode). Alternatively, participants completing interval-contingent assessments reflect on the occurrence of events that took place over the course of a given period at a specified time (e.g., before bed, upon waking, etc.). Lastly, signal-contingent methods rely on a beeper or other electronic device, which allows the research team to send participants a signal prompting them to complete their assessment at that given time. (For a review of diary methods in family research, see Laurenceau and Bolger 2005).

Diary data is valuable for assessing reciprocal influences in family functioning. Larson and Almeida (1999) demonstrate how diary data can be used to examine the ways in which daily interchanges between family members lead to the transmission of emotions. In these instances, repeated diary and experience sampling data shed light on the ways that families function by allowing researchers to map the paths of transmission of emotions in the daily interchanges between family members by capturing the occurrence of events or emotions in one family member and tracking the processes by which subsequent emotions or behaviors occur in other family members. Feeney (2002) capitalized on strengths of diary data to examine the role of romantic attachment in relation to behavior between spouses and marital satisfaction. Through the use of diary data, a more objective description of participants' behavior could be obtained because subjects were not forced to make subjective evaluations of their behavior or summarize broad patterns in their interactions.

Cummings and colleagues have made multiple recent advances in the use of diary methodologies to assess marital conflict in the home and children's responses to it. Based on daily diary data, Cummings et al. (2003) showed the interconnectedness of emotions and behaviors among family members in the context of marital conflict in the home. Analyses revealed compelling evidence for the impact of parents' expression of negative emotions during marital conflict on children's concern, negative emotionality, and insecure behaviors. Cummings et al. (2002) reported in regard to the impact of interparental positive and negative emotionality, respectively, on children's responses to marital conflict, addressing a gap in methodological approaches to studying children's responses to emotion processes associated with interparental relations in the home. Diary data provides a record of children's responses to martial conflict with particularly high ecological validity. In another study, diary data provided a valuable record of the impact of everyday marital conflict on children's aggressive tendencies; furthermore, analogue methods used in the laboratory setting showed consistency with daily diary reports, confirming the reliability of the diary methodology (Cummings et al. 2004).

Recently, Cummings and colleagues have moved in further innovative directions in the use of daily diaries for assessing the effectiveness of psycho-educational prevention programs. In one application, family members (mothers, fathers, and teens) completed daily diaries about family interactions including constructive and destructive tactics during marital and parent-child conflicts (i.e., mother and adolescent, father and adolescent), over the 28 day course of a prevention program aimed at increasing the use of constructive conflict and communication in multiple family systems (Cummings and Schatz 2012). This approach not only assesses the efficacy of the program but is also intended to capture the dynamic patterns of interactions among family members, including the potential to chart the impact of the program on family interactions during the course of the program. Advanced statistical techniques, such as dynamic systems analyses, have enhanced the value of diary data. For example, Schermerhorn et al. (2010) showed how this approach may illuminate complex patterns of bidirectional influence between children and their parents in the context of marital conflict.

Daily diaries provide rich descriptive data that can be used to advance our understanding of the complex and mutually influential nature of family processes. Diaries also may enhance the external validity of laboratory findings. For example, Goeke-Morey et al. (2007) and Cummings et al. (2004) provide demonstrations regarding how findings from analog and diary methods can be mutually supportive in these instances with regard to the impact of marital conflict resolution and interparental aggression, respectively. Because data are not collected in a laboratory setting, however, missing data and the integrity of diary data can be an issue. For example, is it difficult and can be time consuming to ensure that an interval-contingent assessment (i.e., diary entry once a day) is completed at the specified time, or that an event-contingent assessment (i.e., diary entry after a target event, such as a conflict) is completed immediately following the targeted event. The use of electronic devices for the completion of diaries represents a valuable step toward reducing these concerns. Additional strategies for assessing family functioning, such as laboratory observations, can complement diary data by maximizing its strengths and compensating for its weaknesses. Our hope is that diary data, used in conjunction with other measures, will be useful in future family systems work to provide specific insight into the nature of conflict, particularly in instances of extreme adversity in the family (e.g., partner aggression and violence and substance abuse). Diary methodology may be especially valuable to gauge the prevalence and impact of such extreme forms of destructive environments, which are less likely to occur in laboratory research settings (Fals-Stewart 2003; Margolin et al. 2009).

Methods for the Study of Family Systems: A New Coding System for Triadic Interactions

Laboratory observations can capture complex family interactions and processes, provide rich data, and allow for more control over the ways in which participants interact (e.g., control over instructions). Another goal of this chapter is to advocate for the use of a new triadic coding system in the study of family systems. A common strategy among family researchers in a laboratory setting is to construct a situation in which family members have the opportunity to participate in a task designed to elicit specific behaviors. An example from our work is a triadic interaction paradigm in which parents and their adolescent child are asked to discuss a topic that is typically difficult or hard to handle in their family, and then to attempt to arrive at some kind of a solution to that problem. This strategy allows for a higher degree of control than diary methods, and increases the likelihood that specific behaviors and interactions occur and can be recorded objectively. Laboratory observations are limited, however, by their time-intensity and the common need for rigorous coding protocols. The development of automated coding systems represents a step toward addressing these concerns, but forces a trade-off between time-efficiency and the richness and accuracy that human judgment provides. Although research on family interactions suggests that participants tend to interact at they normally would, laboratory observations are not necessarily generalizable to naturalistic settings. This illuminates the value in collecting both naturalistic (diary) observational data in conjunction with laboratory observational data whenever possible (Cummings et al. 2011).

Various coding systems have been established to identify and quantify meaningful patterns of behaviors that emerge over the course of interactions between marital partners, between a parent and a child, and between family members. Some systems allow for coding to occur in the field, as interactions unfold, while others require review of observational records. In either case, coding systems focus on specific constructs of interest and may assess behaviors at either the micro- or macro-analytic level of analysis (Kerig and Lindahl 2001). Micro-analytic systems focus on behaviors as they occur in small time intervals (e.g., 30 or 60 s intervals). This kind of analytic strategy sheds light on the dynamic interplay between participants in an interaction. Macro-analytic systems provide a more global perspective on the interaction as a whole; identifying themes that emerge over the course of the interaction and allowing researchers to discover patterns that may be missed at the micro-analytic level. (For more information on specific coding systems available for family research, see Kerig and Lindahl 2001).

Our research team is currently active in establishing a coding system that integrates the micro- and macro-analytic systems by establishing a series of both global and interval codes that assess instances of specific constructive and destructive behaviors in the context of a family conflict. At the micro-analytic level, triadic family interactions are coded in 30-second intervals on a five-point scale, assessing a range of conflict behaviors from verbal and nonverbal anger, defensiveness, sarcasm and personal insults, to the use of humor, problem-solving strategies, supportive behaviors, and physical affection. Coders make three passes, focusing on each individual participating in the interaction separately. A challenge our team has wrestled with in developing this part of the system is how to capture directionality in these interactions. While the richness of the information obtained is increased by coding to whom the specific behaviors are directed, the process of including this information also dramatically complicates the coding record and dataset. However, the benefit outweighs the cost in our view: Considering directionality in the interval codes provides a more comprehensive picture of the mutually influential transactions that occur between family members. It also allows us to explore dynamics between specific dyads and triads in these family interactions. At the macro-analytic level, global codes complement the interval codes by indicating the "big picture" about family interactions. Specifically, we include codes for the overall degree of autonomy and relatedness between parents and their child, the overall levels of constructiveness and destructiveness in the interaction, positive and negative parenting, and the degree of resolution that the family has succeeded in achieving over the course of their interaction. The global component of this coding system thus allows themes to be characterized that might be missed with an exclusive focus on specific patterns of interaction that occur in 30-second intervals. Taken together, the complementary components of this system provide a means for capturing and understanding transactional family dynamics as they occur over the course of an interaction. In addition to using this system to better understand the complex pattern of interaction between family members from a basic research perspective, preliminary analyses using this system have assessed family-wide outcomes of randomized clinical trials with promising results (Cummings and Schatz 2012).

An additional benefit of laboratory observations that utilize a triadic interaction paradigm is the ability to construct and observe a stressful situation and measure the subsequent physiological response. This provides valuable insight into the characteristics of conflict that are associated with different patterns of physiological responding. Physiological measures allow for the inclusion of reactivity that may not be evident from behavior or self-report. Physiological reactivity can be gauged through the collection of diagnostic fluids such as saliva or blood, or through the use of sensors and imaging devices that measure indicators of physiologic response such as heart rate, blood pressure, and brain wave activity. Physiological measures provide insight into the biological mechanisms by which the environment affects individuals, and may reveal responses to situations or environments that participants are either unable or unwilling to articulate or display through their behavior. Physiological measures are limited in that they can be expensive and time-consuming, not just in terms of collection but also in terms of analysis, often requiring specialized training, equipment, and supplies or external contracts to properly assay samples. Meaningful results may also be hard to obtain or interpret. Finally, participants may be reluctant to participate in projects that include physiological measures, which can be quite intrusive depending on the method of collection and the kind of physiological information gathered. Fortunately, though, there exists a large range of collection techniques, and valuable information can be gained from methods that are minimally invasive.

With regard to physiological reactivity, in the context of marital conflict, atypical patterns of reactivity have been associated with destructive conflict properties. Studies have linked exposure to destructive marital conflict to both internalizing and externalizing problems in children and adolescents. Hypo- and hyper-arousal of the physiological stress response has been implicated in these relationships. There is also evidence for reciprocity in terms of physiologically responses. That is, synchrony and asynchrony in the physiological response of family members during an observed interaction may provide clues into the characteristics and impact of the conflict (Granger et al. 1998). In our own work, a goal for the future is to examine the predictive capacity of reciprocal patterns of physiological responding between mothers and their children over time. In addition, we are beginning to examine relations between triadic family interactions and children's patterns of physiological reactivity to family stress.

Conclusion

The development of new methods holds potential to contribute to new advances in understanding of the impact of family processes and systems on child development. We have shown in this chapter how several directions in methodology may advance understanding of family functioning at a level that reflects the interplay of multiple family systems, both based on methods that may be used in the home (i.e., daily diaries) and laboratory (i.e., triadic coding) and derived from well-constructed analog scenarios for eliciting children's representations of multiple family relationships (i.e., MSSB-R; SSBA). It makes sense for future directions in the development of methodologies to be guided by the increasing interest in more effectively studying mediating and moderating processes in the context of family systems. Relatedly, there is impetus for methodologies that allow us to better examine risk and resiliency, and vulnerability and buffering processes, which may be productively informed by constructs emerging from the developmental psychopathology perspective. For example, patterns of triadic interactions or representations of family relations may inform understanding of children's risk for adjustment problems. Future methodologies may work towards the assessment of physiological responding in the context of family processes and interactions (see Appendix D, Cummings and Davies 2010). Another goal for the future is to tailor methodologies toward further articulating transactional models for interrelations between family processes and child adjustment over time (e.g., Cummings et al. 2008; Schermerhorn and Cummings 2008; Schermerhorn et al. 2008). Processes identified in empirical research concerning family systems' influences on children's risk for the development of psychopathology may also have implications for prevention and intervention research. In conclusion, although theory provides an essential foundation for further advances in understanding, similarly, creative new methodologies can also serve as powerful foundations for new advances. In concert with the development of theory, and informed by integrative reviews of empirical evidence, new methodological approaches to the study of family systems offer the potential for more cohesive, dynamic, and comprehensive depictions of the family as influences on both normal development and the development of psychopathology.

References

Bowlby, J. (1973). *Attachment and loss: Vol. 2. Separation: Anxiety and anger*. New York: Basic Books.

Bradbury, T. N., & Karney, B. R. (1993). Longitudinal study of marital interaction and dysfunction: Review and analysis. *Clinical Psychology Review, 13*(1), 15–27.

Bretherton, I. (1985). Attachment theory: Retrospect and prospect. In I. Bretherton & E. Waters (Eds.), *Growing points in attachment theory and research. Monographs of the Society for Research in Child Development*, 50 (1–2, Serial No. 209), 3–35.

Bretherton, I., & Munholland, K. A. (2008). Internal working models in attachment relationships: Elaborating a central construct in attachment theory. In I. Bretherton, & K. A. Munholland (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (2nd ed, pp. 102–127). New York, NY: Guilford Press.

Bretherton, I., Oppenheim, D., Buchsbaum, H., Emde, R. N., & The MacArthur Narrative Group (1990). The MacArthur Story Stem Battery (MSSB). Unpublished manuscript, University of Wisconsin-Madison.

Bretherton, I., Page, T., Gullon-Rivera, A., Lenzlinger, V., & Munholland, K. A. (2008). Das thema scheidung in den bindungsgeschichten von vorschulkindern: Bedeutung für therapeutische interventionen nach der scheidung (Divorce themes in the attachment stories of preschoolers: Implications for therapeutic interventions with postdivorce families). In K. H. Brisch, & T. Hellbrügge (Eds.), *Wege zur sicheren bindungen in familie und gesellschaft: Prävention, begleitung, beratung, und psychotherapie (Ways towards secure attachment in family and society: Prevention, intervention, counseling, and psychotherapy)* (pp. 279–314). Stuttgart, Germany: Klett-Kotta.

Cicchetti, D., Cummings, E. M., Greenberg, M. T., & Marvin, R. S. (1990). An organizational perspective on attachment beyond infancy: Implications for theory, measurement, and research. In M. T. Greenberg, D. Cicchetti, & E. M. Cummings (Eds.), *Attachment in the preschool years: Theory, research, and intervention* (pp. 3–49). Chicago: University of Chicago Press.

Cox, M. J., & Paley, B. (1997). Families as systems. *Annual Review of Psychology, 48,* 243–267.

Cummings, E. M., & Davies, P. T. (1994). *Children and marital conflict: The impact of family dispute and resolution*. New York, NY: Guilford Press.

Cummings, E. M., & Davies, P. T. (2010). *Marital conflict and children: An emotional security perspective*. New York, NY: The Guilford Press.

Cummings, E. M., & Schatz, J. N. (2012). Family conflict, emotional security, and child development: Translating research findings into a prevention program for community families. *Clinical Child and Family Psychology Review, 15*(1), 14–27.

Cummings, E. M., Davies, P. T., & Campbell, S. B. (2000). *Developmental psychopathology and family processes: Theory, research, and clinical implications*. New York, NY: The Guilford Press.

Cummings, E. M., Davies, P. T., Goeke-Morey, M., & Shamir, H. (2001). Revised MacArthur Story Stem Battery and Coding System. Unpublished manuscript.

Cummings, E. M., Goeke-Morey, M. C., Papp, L. M., & Dukewich, T. L. (2002). Children's responses to mothers' and fathers' emotionality and tactics in marital conflict in the home. *Journal of Family Psychology, 16*(4), 478–492.

Cummings, E. M., Goeke-Morey, M. C., & Papp, L. M. (2003). A family-wide model for the role of emotion in family functioning. *Marriage & Family Review, 34*(1/2), 13–34.

Cummings, E. M., Goeke-Morey, M. C., & Papp, L. M. (2004). Everyday marital conflict and child aggression. *Journal of Abnormal Child Psychology, 32*(2), 191–202.

Cummings, E. M., Schermerhorn, A. C., Keller, P. S., & Davies, P. T. (2008). Parental depressive symptoms, children's representations of family relationships, and child adjustment. *Social Development, 17*(2), 278–305. doi: 10.1111/j.1467-9507.2007.00425.x.

Cummings, E. M., Koss, K. J., & Bergman, K. N. (2011). Assessment of family functioning. In M. El-Sheikh (Ed.), *Sleep and development: Familial and socio-cultural considerations* (pp. 329–354). New York, NY: Oxford University Press.

Davies, P. T., & Cicchetti, D. (2004). Toward an integration of family systems and developmental psychopathology approaches. *Development and Psychopathology, 16,* 477–481. doi:10.1017/S0954.

Davies, P. T., Sturge-Apple, M. L., Winter, M. A., Cummings, E. M., & Farrell, D. (2006). Child adaptational development in contexts of interparental conflict over time. *Child Development, 77*(1), 218–233.

Davies, P. T., Woitach, M. J., Winter, M. A., & Cummings, E. M. (2008). Children's insecure representations of the interparental relationship and their school adjustment: The mediating role of attention difficulties. *Child Development, 79*(5), 1570–1582.

Du Rocher Schudlich, T., Shamir, H., & Cummings, E. M. (2004). Marital conflict, children's representations of family relationships, and children's dispositions toward peer conflict strategies. *Social Development, 13*(2), 171–192.

Fals-Stewart, W. (2003). The occurrence of partner physical aggression on days of alcohol consumption: A longitudinal diary study. *Journal of Consulting & Clinical Psychology, 71*(1), 41–52.

Feeney, J. A. (2002). Attachment, marital interaction, and relationship satisfaction: A diary study. *Personal Relationships, 9*(1), 39–55.

Fiese, B. H., Wilder, J., & Bickham, N. L. (2000). Family context in developmental psychopathology. In A. J. Sameroff, M. Lewis, & S. M. Miller (Eds.), *Handbook of developmental psychopathology* (2nd ed, pp. 115–134). New York, NY: Springer.

Goeke-Morey, M. C., Cummings, E. M., & Papp, L. M. (2007). Children and marital conflict resolution: Implications for emotional security and adjustment. *Journal of Family Psychology, 21*(4), 744–753. doi:10.1037/0893–3200.21.4.744.

Granger, D. A., Serbin, L. A., Schwartzman, A., Lehoux, P., Cooperman, J., & Ikeda, S. (1998). Children's salivary cortisol, internalising behaviour problems, and family environment: Results from the Concordia Longitudinal Risk Project. *International Journal of Behavioral Development, 22*(4), 707–728.

Kerig, P. K., & Lindahl, K. M. (2001). *Family observational coding systems: Resources for systemic research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lamb, M. E. (2010). How do fathers influence children's development? Let me count the ways. In M. E. Lamb (Ed.), *The role of the father in child development* (5th ed., pp. 1–26). Hoboken, NJ: John Wiley & Sons Inc.

Larson, R. W., & Almeida, D. M. (1999). Emotional transmission in the daily lives of families: A new paradigm for studying family process. *Journal of Marriage & Family, 61*(1), 5–20.

Laurenceau, J. P., & Bolger, N. (2005). Using diary methods to study marital and family processes. *Journal of Family Psychology, 19*(1), 86–97.

Margolin, G., Vickerman, K. A., Ramos, M. C., Serrano, S. D., Gordis, E. B., Iturralde, E., & Spies, L. A. (2009). Youth exposed to violence: Stability, co-occurrence, and context. *Clinical Child & Family Psychology Review, 12*(1), 39–54.

Minuchin, P. (1985). Families and individual development: Provocations from the field of family therapy. *Child Development, 56,* 289–302.

Owen, M. T., & Cox, M. J. (1997). Marital conflict and the development of infant-parent attachment relationships. *Journal of Family Psychology, 11,* 152–164.

Rothbaum, F., Rosen, K., Ujie, T., & Uchida, N. (2002). Family systems theory, attachment theory, and culture. *Family Process, 41,* 328–350.

Schermerhorn, A. C., & Cummings, E. M. (2008). Transactional family dynamics: A new framework for conceptualizing family influence processes. In A. C. Schermerhorn & E. M. Cummings (Eds.), Advances in child development and behavior (Vol. 36, pp. 187–250). San Diego, CA: Elsevier Academic Press.

Schermerhorn, A. C., Chow, S., & Cummings, E. M. (2010). Developmental family processes and interparental conflict: Patterns of microlevel influences. *Developmental Psychology, 46*(4), 869–885.

Schermerhorn, A. C., Cummings, E. M., & Davies, P. T. (2005). Children's perceived agency in the context of marital conflict: Relations with marital conflict over time. *Merrill-Palmer Quarterly, 51*(2), 121–144.

Schermerhorn, A. C., Cummings, E. M., DeCarlo, C. A., & Davies, P. T. (2007). Children's influence in the marital relationship. *Journal of Family Psychology, 21*(2), 259–269. doi:10.1037/0893–3200.21.2.259.

Schermerhorn, A. C., Cummings, E. M., & Davies, P. T. (2008). Children's representations of multiple family relationships: Organizational structure and development in early childhood. *Journal of Family Psychology, 22*(1), 89–101. doi: 10.1037/0893–3200.22.1.89.

Shamir, H., Du Rocher Schudlich, T., & Cummings, E. M. (2001). Marital conflict, parenting styles, and children's representations of family relationships. Parenting: Science and Practice, 1(1–2), 123–151.

Sturge-Apple, M. L., Davies, P. T., Winter, M. A., Cummings, E. M., & Schermerhorn, A. C. (2008). Interparental conflict and children's school adjustment: The explanatory role of children's internal representations of interparental and parent-child relationships. *Developmental Psychology, 44*(6), 1678–1690. doi:10.1037/a0013857.

Whiteman, S. D., McHale, S. M., & Crouter, A. C. (2010). Family relationships from adolescence to early adulthood: Changes in the family system following firstborns' leaving home. Journal of Research on. *Adolescence, 21*(2), 461–472. doi:10.1111/j.1532–7795.2010.00683.x.

Chapter 7
Families as Systems: Some Thoughts on Methods and Theory

Robert E. Emery

Systems concepts have had a profound influence across basic and applied sciences notably in biology, engineering, medicine, business, and psychology. Ecology offers perhaps the most widely familiar application of the systems concepts of holism and interdependency. For example, school children are taught that an increase in air or water pollution, or declining acreage in the Brazilian rain forest, can have profound influences on the environment, including on plant and animal life far distant from where the damage occurs. Even young children are introduced to systems concepts in movies like *The Lion King* and its overriding theme, "the circle of life." Linear causality begone!

Today, systems concepts are widely discussed throughout psychology and other social sciences, perhaps nowhere more so than in respect to families. Family systems theory is a catch-all name for a variety of clinical perspectives that date to the 1950s and 1960s, all sharing the common theme of viewing the family as a system. Various applications of concepts like holism, reciprocal causality, boundaries, feedback loops, and subsystems are found throughout family systems theorizing (Minuchin 1985). Unfortunately, family systems theory remains a largely untested clinical approach that includes an amalgam of sometimes compelling hypotheses but no overarching theory other than a general application of systems theory.

Despite limited empirical testing, family systems theory offers a conceptually compelling example of one key tenet offered by systems theory, that is, a rethinking of how systems work. A particularly important aspect of this challenge is the inadequacy of our conceptualizations of simple cause-effect relationships, as embodied by the classic scientific experiment (where there is a clear cause and equally clear effect). Instead, causality in families is reciprocal and interdependent from the perspective of family systems theory. For example, consider the question: Does parental discipline cause children's behavior to change, or does children's (mis)behavior cause parents to exercise discipline? A traditional experimental approach focuses on either parents'

R. E. Emery (✉)
Department of Psychology, Center for Children, Families & the Law,
University of Virginia, Charlottesville, VA, USA
e-mail: ree@eservices.virginia.edu

S. M. McHale et al. (eds.), *Emerging Methods in Family Research,*
National Symposium on Family Issues 4, DOI 10.1007/978-3-319-01562-0_7,

effects on children or children's effects on parents. In contrast, a systems approach sees parental and child behavior as interdependent and influence as reciprocal.

A second central tenet of systems theory is the idea of understanding the same phenomenon at different levels of analysis, an issue that is probably most widely familiar to psychologists in Bronfenbrenner's (1977) ecological theory of human development. Essentially, Bronfenbrenner (1977) reminded psychologists of the importance of the other social sciences by calling attention to nested systems for understanding behavior (the microsystem, mesosystem, exosystem, and marcosystem), as well as potential reciprocal influences across levels.

The themes of nested systems and reciprocal causality are directly relevant to the chapters by Gauthier and Moody (Chap. 5) and Cummings, Bergman, and Kuznicki (Chap. 6) that I comment upon shortly. Before turning to that task, let me be clear that I view systems theory as a broad perspective that scientists can and should use to develop hypotheses, guide innovations in research methods, and understand and integrate empirical findings. I do not expect any study or set of studies to offer some penultimate test of systems theory. In fact, concepts like reciprocal causality cannot be directly tested using the ultimate scientific tool, the experiment. This is not a problem. For example, it is useful to manipulate acid levels in water to observe effects on plant and animal life or to alter parental behavior to observe effects on children. The experiment is not the problem from a systems perspective. The problem is conceptualizing the interdependent, natural world as if it could be divided into independent and dependent variables.

Anatomies of Kinship

In their chapter *Anatomies of Kinship*, Gauthier and Moody (Chap. 5) take a strictly empirical, inductive approach in addressing two questions of huge importance. First, what are the various definitions and forms of families in the US today? Second, what defines various family roles in terms of the activities (role sets) that family and nonfamily members share?

In offering an answer to the first question, Gauthier and Moody (Chap. 5) use cluster analysis to detect the various key configurations that can be observed from household rosters obtained from the General Social Survey. In so doing, they remind us that "family" really should be a plural word, since people in the US today certainly are pluralistic in the families we form. This is an essential reminder, as is the related nested systems idea that family forms have evolved across the course of history in response to broad economic and cultural influences (Hernandez 1993).

Of course, family forms also shape culture and economics in addition to being shaped by these broad social forces. The role of culture in shaping families is evident in the US in longstanding "culture wars" over what constitutes appropriate family roles and what defines "the" family (e.g., a marriage between a man and a woman). The central role of economics in defining "the" family is evident in the welfare reform act of 1996 which begins with the Congressional finding that, "Marriage is the foundation of a successful society," and goes on to explain how the decline in marriage was responsible for the rise in welfare costs in the later part of the twentieth century (Public Law 104–193 1996). At the core, many contemporary political and social debates about marriage and family life concern differences of opinion about whether traditional family functions like housing (welfare), feeding (food stamps), protecting (abuse and neglect), and educating (home schooling) children should be left to parents, or whether these functions should be performed by the state.

Irrespective of one's personal or political viewpoint, Gauthier and Moody (Chap. 5) remind us that, empirically, the idea of "the" family is a myth today. While finding great value in this reminder and also valuing their empirical approach, I simultaneously question the purely inductive method. To begin with, Gauthier and Moody are not studying families. They are classifying households, a fact that they readily acknowledge even though they sometimes refer to their analysis as one of families and other times as one of households. This observation leads immediately to my own frustration with a purely inductive, empirical approach to discovering family forms (or the structure of personality, a favorite occupation of inductive, empirical psychologists (e.g., Zuckerman et al. 1993)). I believe we need theory to guide us. Specifically with respect to Gauthier and Moody's analysis, my question is: When does a household become a family? I doubt that many individuals among the large percentage (17.21 %) of single male households without children revealed by their analysis would define themselves as a family. And I wonder when a cohabiting couple thinks they have become a family. On the day they move in together? When the couple makes a long-term commitment to each other? When they have a baby together? Finally, what about same-sex families? They exist, even though this inductive analysis did not discover them.

Families, it seems, are defined not only by household composition, but also by emotion, personal beliefs, and long-term commitments, particularly to children. This raises another interesting question that Gauthier and Moody didn't address: How do households evolve over time? We know that many cohabiting couples eventually marry; we know many split up (Cherlin 2009). Similarly, we can safely predict that many single men move into cohabitation, marriage, and parenthood (not necessarily in that order). I would love to have these investigators give me the numbers, unfettered by theory, not only in cross-section but also longitudinally.

Let me offer one more thought on this topic before moving to their analysis of role sets. Because people define their family based on more than household composition, families can extend across households, another idea consistent with systems theory. In fact, we know that children who live primarily with their mothers are likely to include their nonresident fathers as members of their family, even though their mothers generally do not (Furstenberg et al. 1983). Of course, this last point means that members of the same "family" may not define their family in the same way. There are psychological families, as well as functional families.

Consideration of function leads into Gauthier and Moody's analysis of role sets. This is an exciting promising approach, defining roles, and perhaps families, more in terms of shared activities than labels. I am not giving them a blank check, however, as I surely would offer my same objections about the importance of self-definition as they further develop this approach. (For example, my own father is very important to me, even though we have no interaction. He is no longer alive. This psychological significance would not be captured by a role sets analysis, which is defined by family interactions.) Actually, what intrigues me most about the potential of their role sets approach is the possibility of identifying *non-normative* family roles. For example, family systems theorists often discuss the "parentified child," a child who is serving more of a parent's than a child's role in a family (Peris and Emery 2005; Peris et al. 2008). Could we use analysis of shared activities to identify such atypical roles? For example, might we define a parentified child as a teenager who assists her mother more than the mother assists her daughter?

Relatively Unstructured Assessment Methods

Although my heading differs from their title, Cummings et al., (Chap. 6) focus much of their chapter on the benefits of using relatively unstructured assessment methods in understanding how parental conflicts affect children. They specifically discuss the MacArthur Story Stem Battery (MSSB), daily diaries, and triadic coding systems, as well as a brief mention of psychophysiological measures. On this last point, let me simply note that I routinely illustrate psychophysiological responses to conflict by screaming loudly at an unsuspecting member of the audience during a lecture. The entire audience is startled and aware of their suddenly increased physiological arousal, an experience that allows us to better appreciate one aspect of a child's dilemma when exposed to loud, angry, and likely prolonged and repeated parental conflict.

I welcome Cummings et al.'s, (Chap. 6) use of less structured measures for several reasons. First, rather than simply imposing investigators' hypotheses on research participants, as is necessarily the case with highly structured measures, their relatively unstructured measures allow us to learn from our participants. For example, we can learn about children's (perhaps exaggerated) fears in response to conflict, children's efforts to comfort or confront parents in conflict (rather than passively responding to them), or about how parents attempt (or fail) to protect their children from their disputes. Second, unstructured measures reveal that children's psychological life is deeper and richer than what we can capture on useful but limited structured measures like the CBCL (Child Behavior Checklist) (Achenbach 1999), a fact that is of great interest to me both as a parent (I worry about far more subtle matters than CBCL items) and as a researcher (e.g., my interest in psychological pain, see Laumann-Billings and Emery 2000). Third, less structured measures show us that conflict is multidimensional including not only unconditioned psychophysiological reactions to anger, but also less effective individual parenting, inconsistent coparenting, and inherent loyalty dilemmas about whether and whom third parties (children) should side with in a dispute. This last issue also underscores the value of triadic coding systems, an effort I have pursued myself (Vuchinich et al. 1988). Recently, I have grown concerned about the effects of extreme loyalty dilemmas, specifically deliberate efforts on the part of some parents in high conflict to denigrate the other parent in an effort to form an alliance with a child against the other parent (Rowen and Emery 2012).

While applauding the expanded reach of their less structured measures, let me also underscore the essential value of Cummings et al. (Chap. 6) continued commitment to reliable, objective coding of these measures. Their assessment methods are more qualitative, but they quantify their participants' responses, even though unstructured information often is quite complicated to score objectively. This time consuming commitment to objective coding of less structured data expands the reach of the work of Cummings et al. in another direction. It gives their studies broad scientific impact, rather than limiting their effort to useful, but limited hypothesis generation.

Finally, Cummings et al.'s (Chap. 6) concern with theory, notably their emotional security hypothesis, is a great example of why I believe theory is necessary to guide scientific understanding. Some early investigations of the effects of parental conflict seem to have viewed children as automatons, who mindlessly imitated misbehavior modeled by their parents. While not discounting the importance of modeling, Cummings' et al. theoretical, systemic approach reminds us that parental conflict can affect children through multiple avenues in multiple domains, in particular by undermining the sense of emotional security that is an essential aspect of healthy child development.

Content and Meaning in Family Interaction

Let me turn to a discussion my own work, and offer the first of two illustrations about the value of systems thinking and the need for theory. As I discussed about two decades ago (Emery 1992), family systems theorists have offered many useful ideas about how families work like systems, but have failed to elaborate about what it is that family members exchange. That is, family systems theorists have emphasized process but often overlooked content. What purposes do families serve? Earlier, I noted several broad functions that families serve including housing, feeding, protecting, and educating children. Here, I turn to more internal, emotional functions of family relationships.

Love and Power

Conceptually—and to a lesser extent, empirically—I suggest that there are two overarching substantive dimensions that form the basis of family relationships and serve critical interpersonal functions: love and power. "Love" in this context refers to the wonderful, ephemeral emotion we all seek, but the term also refers more generally to a dimension of emotional involvement. Very involved relationships that often but not always are loving, anchor one end of the continuum, and emotionally distant businesslike relationship are at the opposite extreme.

"Power" refers to dominance, a term that should not be confused with dominating, because one does not have to exercise power in order to have it. Like the ethological concept of dominance, I define power by the privileges that accompany an achieved status. Parents normatively have power over children in a family, and they retain their power (usually) even when they let their children "get away with murder." "Let" is the key concept, as the term implies that a parent can retain their power even when they choose not to exercise it.

What functions does love serve in a family? I suggest there are several functions (and types) of love, including promoting security, encouraging identity, building loyalty, strengthening pair bonding, sexual expression, and committed parenting. For infants, love is attachment and the security that grows out of responsive parenting. For older children, love fosters identity and the adoption of parental and family values. For siblings, and parents and children, love promotes loyalty–support when help is needed, an allegiance when under attack. Pair bonding, sexuality, and committed parenting all are a part of healthy love between parents, although each function is probably separate as is observed across animal species and within human individual differences.

What is the function of power? Again, there are several closely related functions of power including the socialization of children, establishing hierarchies that reduce conflict (e.g., in sibling relationships or parent-child relationships), and the efficiencies that come with role specialization. On the last point, power is not necessarily all encompassing, as for example, one spouse might hold more power over finances and the other more power over relationships. And while parents normatively hold power over children, the course of child development involves parents gradually ceding power to children. Once authority over a specific issue is given up (e.g., a child is allowed to go out alone), a parent's legitimate authority to control that particular issue is pretty much lost, but parents nevertheless retain their dominance role. This role includes the right to revoke privileges that are abused or as a punishment (e.g., you can no longer go out alone).

I offer little direct defense here of love and power as the two basic dimensions of family relationships. Instead, I hope to make a case for the value of these two dimensions by illustrating how the ideas can lead to a deeper understanding of families, particularly family conflict. I do note however, that my focus on love and power is not unique, an observation that I hope lends credibility to my argument. Dominance and affiliation are basic dimensions used by ethologists to describe animal relationships (Omark 1980). A metaphorical view can reinterpret Freud's basic instincts of sex and aggression more broadly as representing love and power. Harry Stack Sullivan's (1953) interpersonal theory of personality, especially as operationalized by the circumplex model, is very clearly constructed around these two dimensions (Leary 1957) (although I believe, for reasons I note shortly, that Sullivan defined the love dimension incorrectly). Finally, developmental psychologists clearly use the two dimensions in creating the widely accepted typology of parenting into four categories: the authoritative (high love, high power), authoritarian (low love, high power), permissive, (high love, low power) and disengaged (low love, low power) parenting styles (Maccoby and Martin 1983).

Surface and Deep Meanings of Conflict

Love and power are evident in a variety of aspects of family relationships, but these qualities are perhaps most clearly revealed during family conflicts. Conflict is revealing, because, according to the theory I put forth here, conflict is a way of asserting, changing, or resisting change in the love and power *boundaries* of a relationship. The term boundaries, refers to the edge of an individual's or a relationship's territory, much like a perch on a high tree can mark the boundary of the territory of a songbird.

I argue, in fact, that all family conflicts have two levels of meaning. The first (surface) level concerns the topic of the dispute, and disputes are wide ranging in families. Siblings fight about who gets to ride "shotgun" in the family car. Children argue with each other, and with their parent, about who gets to sit in the parent's lap first. Spouses argue about getting home late and whether the partner who is late should have called to warn their spouse about the delay. And, of course, the list could and does go on and on.

At the *surface* level of analysis, I suggest that these kinds of disputes are about sitting shotgun, going first, and being late. At a second *deep* level of meaning, however, I argue that these disputes are about much more: Love or power or both. Getting to ride shotgun is not a privilege to be shared. It is a mark of rank, of power, that goes to the oldest child, even when challenged, even (or especially) when a younger sibling "got there first." Being picked first is a sign of a parent's preference (unless carefully balanced over time), with the preferred child being favored, a meta-communication about love that siblings often taunt each other about following a victory. ("See? Mom loves me more!") For one spouse, being late unapologetically may be a means of exercising or reaffirming their power ("You can't control me"), as might their refusal to call to give notice ("You're not my mother"). To the other spouse, however, the tardiness conveys an absence of caring and concern (a love struggle), especially when combined with lack of consideration of a telephone call.

In short, much like a battle between two bucks over the opportunity to mate with a cow is about more than sex (i.e., dominance), I suggest that family conflicts have both surface and deep meanings. At the deep level of analysis, family conflicts are either power struggles or love struggles or some combination of both.

Dominance is a familiar example of the deep meaning of power struggles, and the very term, power struggle, connotes the more basic (deep) implication of the outcome of conflict. Power is defined and attained by winning or more generally by having control over conflict resolution. As noted earlier, a parent who "allows" their child to win a dispute, perhaps accepting a later curfew, retains power over the child (although perhaps no longer in relation to the particular topic of dispute). And in triadic interaction, a parent (or judge) who decides the outcome of a conflict retains power over the disputants even though the third party has no stake in winning.

What about love struggles? The term is unfamiliar and awkward, perhaps because love struggles are not so obvious to us. There also is no clear animal analogue for love struggles. What is the goal of a love struggle? Unlike a power struggle, I suggest that winning carries little meaning at a deep level. Instead, the expression of emotion is the key signal sought in a love struggle, for example: "I'm so sorry. I didn't mean to hurt your feelings. I should have called." In fact, the valence of the emotional expression matters less than its intensity. The test of emotional investment in a love struggle is the expression of strong emotions. Thus, when a married spouse screams, "I want a divorce," because they are feeling hurt and angry (i.e., they don't really want a divorce; they want reassurance of their partner's love), their partner shows their own emotional investment (the key to a love struggle) whether they cry and bellow, "I couldn't live without you" or if they scream back, "If that's what you want, go for it! But, baby, you are going to pay!" Contrast the intensity of these reactions of opposite valence and the emotional investment each conveys, with a calm, "OK. If that's what you want." This calm detached remark conveys emotional distance, and it is not at all what the spouse who is screaming about divorce is looking for. (A partner who really wants a divorce typically has a plan, much like someone who is serious about suicide has a plan (Emery and Sbara 2002)).

This leads to an important point that often is misunderstood by disputing couples and by psychologists. Love and hate are not far apart. Both convey a high degree of emotional investment (or what I have called "love"). Lovers, divorced spouses, or parents and children who fight constantly and intensely are very much involved with each other. The opposite of love is not hate (as Sullivan (1953) asserted, I think incorrectly). The opposite of love is indifference.

Love Struggles and Power Struggles in Families: Some Examples

Having outlined a handful of basic assumptions, let me offer some examples of how we can see family conflicts in a different light if we view them as being either love struggles, power struggles, or both at a deeper level of meaning. Rather than marshaling data and studies, in the present context these examples will stand as the test of these ideas.

Changing Boundaries of Parental Discipline

Viewing parental discipline in terms of dominance relations is one interesting reconceptualization suggested by the power struggle/love struggle analysis. Other than the literature on parenting styles, most writing about parental discipline focuses at the microanalytic level, for example, considering what behaviors parents should or should not tolerate and how parents should punish, reward, or ignore their children's behavior (in addition to loving them). Developmental themes are evident in the relevant professional and popular literature, but there is no underlying explanation of developmental change. I suggest a power struggle/dominance reconceptualization of effective parental discipline based on the principles that (1) parents should remain dominant over their children throughout childhood (dominant is not dominating), (2) the boundaries of parental authority shrink and the boundaries of children's autonomy expand as children grow older, (3) the most important conflicts over discipline occur at these boundaries, because a change in the "winner" leads to a change in the structure of the relationship (i.e., the child has expanded her boundaries by gaining a new area of autonomy), and (4) consistent parental love is essential to effective discipline/dominance for many reasons, not the least important of which is so that love struggles and power struggles do no get confused (see point 3 in discussion of coercion below).

Together, these principles indicate that the key to maintaining effective parental discipline is to remain in a position where a parent *can* exercise control when needed, but that authority is used judiciously based on the child's age and various circumstances. The ideal in fact, is for parents to use discipline rarely, because the boundaries of their authority are clear. In everyday terms, this means that children know what they can and cannot do.

Children's growing sense of autonomy is the counterpart to parental authority. Within their boundaries of autonomy, children pretty much get to make their own decisions, albeit with a degree of parental oversight. Occasionally, parents resume authority deep within a child's territory of autonomy (e.g., they revoke privileges or veto wearing inappropriate clothing to school), and of course, children sometimes assume authority for things that clearly lie within their parents' realm (e.g., sneaking drinks from the liquor cabinet). But the really important and intense conflicts are boundary conflicts, those grey areas of authority/autonomy (e.g., "All my friends can go to night games alone!") that change rapidly around developmental transitions. Parents and children must negotiate carefully in regard to boundary conflicts, both because once autonomy is gained it is hard to wrest it away and because exercising a degree of autonomy offers children important practice and contributes to children's growing sense of self.

Anecdotally, I can say that the dominance reconceptualization has caused me to think and feel entirely differently about parenting my own five children. I focus more on the big picture (being Mufasa, the original Lion King) and more readily cede my position in minor disputes (which are not really tests of my authority). And I can say I particularly feel the dominance competition when playing sports with my boys. Winning (or increasingly, losing) at ping pong, soccer, or basketball is a test of dominance, not just a game with my sons. A victory underscores my dominance. A loss signals one more victory for them in their march to unseat me. Still, I usually remind them, "Even though you won this time, Pride Rock is not yours yet, Simba."

Coercion

A reconceptualization of the very useful behavioral family therapy construct of coercion (Patterson 1982; Forgatch and Patterson 1998) is perhaps my favorite example of applying the power struggles/love struggles framework. The coercion construct views dysfunctional parent-child interactions as resulting from "reinforcement traps," where parents unwittingly reward children's misbehavior. The classic example is when parents give in to escalating child misbehavior.

I want an ice cream.
No. It's dinner time.
I WANT AN ICE CREAM!
I said NO!
GIMME AN ICE CREAM! GIMME AN ICE CREAM! I HATE YOU!! GIVE ME AN ICE CREAM!
OK. Be quiet. But you better eat all your dinner.

In this example, the parent accidentally gives the child positive reinforcement (the ice cream) for increasingly obnoxious behavior. Importantly, the coercion construct views the interaction in systems terms by noting that the child also negatively reinforces the parent by ceasing their obnoxious behavior once the parent capitulates (reciprocal causality).

A related behavioral family concept is "negative attention seeking," where an action that the parent views as a punishment (e.g., yelling at a child) functionally serves as reinforcement. (i.e., When the parent yells, the obnoxious behavior increases rather than decreases). From an operant conditioning perspective, the "punishment" is really a reinforcer.

Let me be clear. I find these behavioral family therapy constructs very useful conceptually and practically, and there is a solid evidence base for treatments based on these principles (Forgatch and Patterson 1998). I suggest, however, that the operant analysis of these constructs is deficient in three ways.

First, I wonder why the following is not considered to be a coercive interaction:

Make your bed.
I don't feel like it.
I DON'T CARE IF YOU DON'T FEEL LIKE IT. MAKE YOUR BED.
I DON'T WANT TO.
Make it right now, or you will be in big trouble.
Arggh. Oh. OK.

From an operant perspective, this interaction is maintained by positive and negative reinforcement, exactly as in the ice cream example. But in this case the reinforcement is not generally viewed as "accidental." Why? The parent wins, whereas the child won in the ice cream example. In short, the uncontroversial idea that parents should have more power than children is implicit in the coercion construct. The power struggle analysis makes this explicit. In fact, some family systems theorists talk about "inverted hierarchies" in families where children have inappropriate power over parents.

The second way in which the power struggle/love struggle analysis adds to the operant understanding of dysfunctional interactions is by adding a developmental perspective on parent-child disputes. Parents of toddlers in throes of the "terrible twos" often do not have the patience to win every battle that results from a temper tantrum. And perhaps parents should not win every time. As suggested by the idea of changing boundaries of parental authority, parents do not want to thwart the 2 year olds growing sense of independence and control. So, while no parent should give into a toddler's tantrum about wanting to play in a busy street, perhaps it is not worth the battle (and harmful to the child's growing sense of control) to force a toddler to eat every healthy food they reject or to insist that the toddler consistently wears a certain item of clothing, when the prospect of doing so sends the toddler into a raging tantrum.

Parents "losing" coercive interactions similarly may not only be acceptable but also beneficial during another stage of rapid development, early adolescence. 14 year olds should have curfews and be monitored in various ways, but early teens probably should win at least some battles over things like cleaning their room, clothing (again), choice of friends, and earning new privileges. This gives the early adolescent a sense of control and early practice in making decisions autonomously, something that parents want to prepare teenagers to do on their own.

The third insight offered by the power struggles/love struggles analysis for the coercion construct is an answer to the questions: Why is the child trying to get an ice cream cone by using obnoxious tactics instead of positive (if perhaps manipulative) ones? Why do some children find "negative attention" to be rewarding rather than punishing? An operant analysis does not attempt to answer these questions. Instead, an operant approach defines reinforcers and punishers by their effects. I suggest however, that some, perhaps many, coercive interactions are not power struggles as the behavior family therapy literature (implicitly) assumes (see point 1 above). Instead, maybe these interactions are love struggles. Negative attention is rewarding because the alternative is no attention. Children use negative tactics, because parents ignore their positive initiations. If this is the case, such interactions really are not about winning. They are instead a test of the parent's emotional investment. The best treatment may not be for parents to learn to win during coercive interactions, but for parents to increase the attention they pay to their children for positive behavior (or noncontingently), so children do not need to seek attention in negative ways.

Love Struggles

More generally, the suggestion that many conflicts in relationships are about love, not power, is one of the most important contributions of the power struggles/love struggles analysis. For me, this insight became must compelling in my research comparing the mediation and litigation of child custody disputes. In my studies, I worked extensively with parents who were locked in an apparent power struggle about which parent should have care and control of their children. I quickly became apparent to me clinically, and in some research (Emery 2011; Sbarra and Emery 2005, 2008) that many divorced parents were not *really* fighting about their children. They were fighting about the end of their relationship. In one-sided divorces, the partner who was left used disputes over the children (or money) in a desperate attempt to hang on to their former spouse. The unspoken and perhaps unconscious tactic was something like, "I'll make leaving so miserable that you won't leave me." The goal wasn't to make the spouse miserable. The goal was to get them to stay.

This angry indirect bid to hang on (or at least garner short-term interest and attention) is evident in all kinds of less complicated interactions. A toddler becomes anxious when he wonders too far from an attachment figure, and if he cannot find his way back, he moves on to the next strategy: an angry scream. “MOMMY!” Such “reunion behavior” (Cassidy and Shaver 2008) in attachment terms is likely the developmental reason why so many love struggles get played out as angry conflicts. I also see love struggles as explaining why fourth and fifth grade boys often spend recess after recess chasing and “torturing” girls because they “hate” them so much. (I think something else may be going on.) Perhaps the dysfunction of the patient with borderline personality disorder involves some chronic misuse of the same strategy, as dramatic angry actions are used repeatedly to test and engage others, a tactic that of course, works poorly when it is the dominant strategy. Finally, love struggles are at the heart of the “pursuer-pursued” relationship (Eldridge and Christensen 2002), where the more invested partner constantly seeks indirect signs of reassurance of their partner’s love (“Maybe we should break up!”). The correct answer to which is an emotionally intense response (“I CAN’T STAND IT WHEN YOU SAY THAT!”) rather than an emotionally detached answer (OK. If that’s what you want . . .”).

Systems of Emotion

I want to offer what I find to be a second, compelling application of systems theory, an example that I can discuss only briefly. This is the growing idea based in affective neuroscience that basic emotions are not isolated feeling states but part of a system of motivations and actions that serve both immediate and evolutionary functions (Panksepp 1998). The neuroscientist, Jaak Panksepp (1998), discusses this idea at length and postulates that there are at least seven basic human/mammalian emotional motivations. I use the term emotional motivations as a way of indicating that Panksepp (1998) views basic emotions as very much akin to motivations like hunger, thirst, and sex. The internal feeling state of hunger, for example, is only a part of a constellation of connected feelings and actions that includes hunger, search, gathering/hunting, consumption, and satiation. He argues that, both from evolutionary and neuroscience perspectives, basic human emotions, which are housed in subcortical structures of the brain (our mammalian and reptilian brain) similarly are not separate but connected to other actions and feeling states.

My goal here is not to defend Panksepp’s affective neuroscience or his list of basic emotions. Instead, I want to offer a few examples of my own theorizing about emotions that I developed primarily as a result of my clinical work and that I think are broadly consistent with affective neuroscience. For present purposes, the more important point is that my views and Panksepp’s (1998) affective neuroscience both conceptualize emotions in systems terms.

Many theories treat emotions in a way that resembles a color wheel. From this perspective, emotions are independent feeling states; they are connected to action only through experience and learning. Like colors on a wheel, some emotions are similar to one another, others are the opposite of one another, and various emotions can be blended to form new shades of feeling.

Why would evolution select emotions in this way? That is a question I cannot answer, but it is a question that I think any theory of emotion must answer. Yet, I can easily (if superficially) understand how evolution could select hunger, thirst, and sexual interest, all of which serve obvious immediate and ultimate (evolutionary) functions. I similarly find it easy to intuitively understand how evolution might select somewhat more complicated behavioral systems like fight or flight, attachment, responding to pain, or even grief. We observe these behavior patterns in other animals. Fight or flight is ubiquitous, as are defensive responses to the pain of physical injury. Some form of infant-caregiver attachment is observed in all animals and some birds. Even elephants (and other mammals) grieve (Moss 2000).

None of this is new. But this leads to some interesting and I think clinically relevant observations as we dissect implications for how emotions work. Consider the example of attachment, which is widely accepted and studied by psychologists and biologists (Bowlby 1969, 1973, 1980; Cassidy and Shaver 2008). The infant's attachment bond begins with an internal feeling state–call that emotion feeling secure, loved. As the infant can begin to wander away from its protective attachment figure (and the beginning of locomotion is the time in development when attachment bonds grow strongest), the infant feels another emotion when it wonders too far–call that emotion anxiety. If reunion is somehow blocked, the infant (or more likely a toddler now) will begin to tantrum and rage for its caregiver–call that emotion anger. Finally, when reunited with the caregiver, the infant feels soothed, comforted, and secure. These internal feeling states together serve both immediate and ultimate evolutionary functions, namely promoting proximity to caregivers which leads to increased survival for vulnerable offspring.

This short summary of the attachment mechanism illustrates my point that love, anxiety, and anger are intimately connected in the attachment system. Together they produce the feelings of security that encourage a developmentally appropriate degree of physical independence and also the anxiety and anger that encourage reunion behavior when the child moves outside of the secure range. Neither love nor anxiety nor anger exists independently. They are a part of a system of emotions.

Things get more interesting as we simultaneously consider other behavioral systems/emotional motivations. Fight or flight involves anxiety and anger. Response to pain involves pain, an emotion that in humans includes social pain, for example, hurt feelings (MacDonald and Leary 2005; Panksepp 2005) and anger. Grief is a bit more complex, but elsewhere I have described grief as involving a cycle of love, anger, and sadness (Emery 2011).

Note that I have just described four different emotional motivations each of which includes the same emotion of anger. Yet, I would suggest that anger is not the same in each system. As I often say, anger is not anger is not anger. By this I mean that anger serves different functions in different contexts depending upon which basic emotional system(s) are engaged. The toddler's angry screams may look and feel the same as an angry defense in the face of threat, or an angry retaliation when we are hurt, or the anger we feel over a loss (which probably is a variant on reunion behavior), but when we look at the context, when we understand the basic emotional motivation, we can understand that indeed, anger is not anger is not anger.

I developed much of my thinking about emotional motivations in my work mediating divorce custody disputes, a circumstance in which emotion is raw and basic and where anger dominates (Emery 2011). These clinical insights allow me to understand and respond to anger in different ways, depending upon which motivational system is activated. Just like we reflexively scream at a piece of furniture when we unexpectedly stub a toe, much anger in mediation is a response to the emotional pain of loss. (And much like we sometimes stupidly kick a chair a second time after stubbing a toe, the pain of a marital separation can also set off a primitive impulse to hurt back.) So rather than responding to anger as an attempt to dominate (another basic behavioral system with its unique constellation of emotions), I am able to get to the true feeling lying behind much anger when I call a "timeout" in mediation and caucus with the hurt partner alone. (As I have described at length, difficult divorces typically are one-sided, and the hurt partner is the "difficult" one in mediation.) My strategy is simple. I will tell the angry/hurt man (and it's usually a man), "You seem angry." But after a quick acknowledgement of that surface feeling, I urge him down a completely different emotional trail by saying, "But I think something else is going on." I cannot count the number of men who burst into tears at this point in mediation, as they share stories that are painful indeed. And their focus on their true, deeper emotion in the pain-anger emotional system allows us to move forward at a time when persistent anger would leave us stuck or worse (Emery 2011).

I could offer similar examples for the other emotional motivations behind anger that I have described. Sometimes the opposite of anger is not pain but love and longing, that is, much of the conflict expressed between divorcing partners can be viewed as a form of reunion behavior. Again, the opposite of love is not hate. It is indifference, and I encourage divorcing couples to move from their overinvolved anger to develop a businesslike coparenting partnership (Emery 2004). In other cases, the opposite of anger is fear; in still others, it is sadness or grief. If so, I urge my clients to acknowledge and address their fear or grief, rather than covering it with anger.

The specifics of anger in mediation do not matter as much as the broader point: A systems perspective not only helps us to understand family relationships, but it also helps us to better understand the emotions that serve to motivate all kinds of functional and dysfunctional behaviors in family relationships. At one level of analysis, the human body is comprised of a variety of systems (e.g., the circulatory system, the digestive system, etc.) that we understand much better when we know their immediate and evolutionary functions. The human brain also is comprised of a collection of systems, with the subcortical brain executing many fairly fixed and specific programs in which internal states motivate behaviors that also serve immediate and evolutionary functions. At a broader level of analysis, individual humans are a part of a variety of interpersonal systems, the most central of which typically is one family form or another.

References

Achenbach, T. M. (1999). The child behavior checklist and related instruments. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment* (2nd ed., pp. 429–466). Mahwah: Lawrence Erlbaum.

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist, 32,* 513–531.

Bowlby, J. (1969). *Attachment and loss: Volume 1: Attachment*. London: Hogarth Press and The Institute of Psychoanalysis.

Bowlby, J. (1973). *Attachment and loss: Volume 2: Separation, anxiety and anger*. New York: Basic Books.

Bowlby, J. (1980). *Attachment and loss: Volume 3: Loss: Sadness and depression*. New York: Basic Books.

Cassidy, J. & Shaver, P. R. (Eds.). (2008). *Handbook of attachment: Theory, research, and clinical applications* (2nd ed.). New York: The Guilford Press.

Cherlin, A. J. (2009). *The marriage go-round: The state of marriage and the family in America today*. New York: Knopf.

Eldridge, K. A., & Christensen, A. (2002). Demand-withdraw communication during couple conflict: A review and analysis. In P. Noller, & J. A. Feeney (Eds.), *Understanding marriage: Developments in the study of couple interaction* (pp. 289–322). New York: Cambridge University Press.

Emery, R. E. (1992). Family conflict and its developmental implications: A conceptual analysis of deep meanings and systemic processes. In C. U. Shantz & W. W. Hartup (Eds.), *Conflict in child and adolescent development* (pp. 270–298). London: Cambridge.

Emery, R. E. (2004). *The truth about children and divorce: Dealing with the emotions so you and your children can thrive*. New York: Viking/Penguin.

Emery, R. E. (2011). *Renegotiating family relationships: Divorce, child custody, and mediation* (2nd ed.). New York: Guilford.

Emery, R. E., & Sbarra, D. A. (2002). What couples' therapists need to know about divorce. In A. S. Gurman, & N. S. Jacobson (Eds.), *Clinical handbook of couple therapy* (3rd ed., pp. 508–532). New York: Guilford.

Forgatch, M. S., & Patterson, G. R. (1998). Behavioral family therapy. In F. M. Dattilio (Ed.), *Case studies in couple and family therapy: Systemic and cognitive perspectives* (pp. 85–107). New York: Guilford.

Furstenberg, F. F., Peterson, J. L., Nord, C. W., & Zill, N. (1983).The life course of children of divorce: Marital disruption and parental contact. *American Sociological Review*, *48*, 656–668.

Hernandez, D. J. (1993). *America's children: Resources from family, government, and the economy*. New York: Russell Sage.

Laumann-Billings, L., & Emery, R. E. (2000). Distress among young adults from divorced families. *Journal of Family Psychology, 14,* 671–687.

Leary, T. (1957). *Interpersonal diagnosis of personality*. New York: Ronald Press.

Maccoby, E. E., & Martin, J. A. (1983). Socialization in the context of the family: Parent–child, interaction. In E. M. Hetherington (Ed.), *Handbook of child psychology* (4th ed., Vol. 4, pp. 1–101). New York: Wiley.

MacDonald, G., & Leary, M. R. (2005). Why does social exclusion hurt? The relationship between social and physical pain. *Psychological Bulletin, 131,* 202–223.

Minuchin, P. (1985). Families and individual development: Provocations from the field of family therapy. *Child Development, 56,* 289–302.

Moss, C. (2000). *Elephant memories: Thirteen years in the life of an elephant family*. Chicago: University of Chicago Press.

Omark, D. R. (1980). The group: A factor or an epiphenomenon in evolution? In D. R. Omark, F. F. Strayer, & D. G. Freedman (Eds.), *Dominance relations: An ethological view of human conflict and social interaction* (pp. 21–67). New York: Garland.

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. London: Oxford University Press.

Panksepp, J. (2005). Why does separation distress hurt? Comment on MacDonald and Leary. *Psychological Bulletin, 131,* 224–230.

Patterson, G. R. (1982). *Coercive family processes*. Eugene: Castalia.

Peris, T., & Emery, R. E. (2005). Redefining the parent-child relationship following divorce: Examining the risk for boundary dissolution. *Journal of Emotional Abuse, 5,* 169–189.

Peris, T. S., Goeke-Morey, M. C., Cummings, E. M., & Emery, R. E. (2008). Marital conflict and support-seeking by parents in adolescence: Empirical support for the parentification construct. *Journal of Family Psychology, 22,* 633–642.

Public Law 104–193, (1996).

Rowen, J., & Emery, R. E. (2012). Parental denigration: An empirical study of high conflict behaviors across marital status and their impact on parent-child relationships in young adulthood. Unpublished manuscript, University of Virginia.

Sbarra, D. S., & Emery, R. E. (2005). Coparenting conflict, nonacceptance, and depression among divorced adults: Results from a 12-year follow-up study of child custody mediation using multiple imputation. *American Journal of Orthopsychiatry, 75,* 63–75.

Sbarra, D. S., & Emery, R. E. (2008). Deeper into divorce: Using actor-partner analyses to explore systemic differences in coparenting following mediation and litigation of custody disputes. *Journal of Family Psychology, 22,* 144–152.

Sullivan, H. S. (1953). *The interpersonal theory of psychiatry*. New York: Norton.

Vuchinich, S., Emery, R. E., & Cassidy, J. (1988). Family members as third parties in dyadic family conflict: Strategies, alliances, and outcomes. *Child Development, 59,* 1293–1302.

Zuckerman, M., Kuhlman, D. M., Joireman, J., Teta, P., & Kraft, M. (1993). A comparison of three structural models for personality: The big three, the big five, and the alternative five. *Journal of Personality and Social Psychology, 65,* 757–768.

Part III
Approaches to Measuring Families

Chapter 8
Studying Family Transitions from a Systems Perspective: The Role of Biomarkers

Carolyn Tucker Halpern, Kathleen Mullan Harris and Eric A. Whitsel

Family processes—and the individual developmental transitions embedded within them—reflect complex interactions among multilevel factors including genetic, hormonal, and neural influences; higher-level cognitive, experiential, and behavioral processes; as well as the physical, social, and cultural environments in which they operate. Family influence and process are by definition, partly biological. Genetic endowments and the shared environment of the family confer both sensitivities and vulnerabilities to family members, as well as developmental opportunities to foster resilience. To understand the complex intersections of these diverse factors, a multilevel dynamic systems approach is needed. By multilevel, we do not intend a statistical definition but rather a conceptualization that encompasses the broad range of factors noted above and their coactional contributions to biological and social processes (Gottlieb and Halpern 2002). Here, we discuss biological factors as contributors to the family system and biomarker collection in large scale studies framed through our experiences in the National Longitudinal Study of Adolescent Health (Add Health). We illustrate field, laboratory, and data dissemination challenges for a selection of common biomarkers, leading to best practice recommendations. We also present illustrative findings from research integrating biomarker, social and behavioral data that provide novel insights into social and behavioral phenomena.

"…. sociological problems are better understood when a biosocial theory is brought to bear."
J. Richard Udry, 1988, page 717

C. T. Halpern (✉)
Department of Maternal & Child Health, University of North Carolina
Chapel Hill, Chapel Hill, NC, USA
e-mail: carolyn_halpern@unc.edu

K. M. Harris
Department of Sociology, University of North Carolina Chapel Hill,
Chapel Hill, NC, USA
e-mail: kathie_harris@unc.edu

E. A. Whitsel
Departments of Epidemiology and Medicine in the Gillings School of Global Public Health
and School of Medicine, University of North Carolina, Chapel Hill, NC, USA
e-mail: eric_whitsel@unc.edu

S. M. McHale et al. (eds.), *Emerging Methods in Family Research,*
National Symposium on Family Issues 4, DOI 10.1007/978-3-319-01562-0_8,

Finally, we offer a rationale for incorporating biomarkers into social science research, despite the challenges, and highlight future possibilities for expanded multilevel research capitalizing on intergenerational study designs.

Why Consider Including Biomarkers?

Family process is inextricably linked to health, and the study of health is fundamentally a study of biological processes and outcomes. Social science and health surveys have traditionally relied on self-report to identify health outcomes. For example, at several waves of data collection, Add Health presented a list of chronic conditions to respondents (Rs) and asked, "Has a doctor, nurse or other health care provider ever told you that you have or had. . . .(cancer or lymphoma or leukemia; high blood cholesterol or lipids; high blood pressure or hypertension; high blood sugar or diabetes; . . .)." Rs self-reported "yes" or "no;" affirmative responses were followed by a question about age of onset.

Although self-reported health measures vary in quality, they generally underestimate health risks that go undetected or for which symptoms do not appear early or consistently; this is especially true among young, otherwise healthy populations (Kehoe et al. 1994). Moreover, socioeconomic status, race, ethnicity, nativity, and language proficiency influence knowledge of health, symptoms, and disease; access to diagnostic services; and understanding of health questionnaires. Consequently, the accuracy of self-reported health survey data can vary in ways unrelated to pathophysiological mechanisms of disease. Objective measures are therefore the gold standard for reliable and valid measurement of health.

In a survey field setting, objective health measures are those derived from the collection of biospecimens or through physical measurement by trained and certified personnel (e.g., interviewers or phlebotomists) following standardized protocols. We refer to these objective measures as "biomarkers" or "biological markers" of (ab)normal biological states resulting from underlying (patho)physiological processes. Biomarkers were once limited to clinical patient samples, or community-based epidemiological studies. However, in the early 1990s several international and/or aging studies began collecting biomarker data to strengthen and complement self-reported data on health and aging (Finch et al. 2001; Weinstein et al. 2008). For example, the Indonesian Family Life Survey (IFLS), the National Study of Midlife Development in the United States (MIDUS), and the Social Environment and Biomarkers of Aging Study (SEBAS, i.e. Taiwan Biomarker Project) were forerunners in this regard. However, these studies focused on later life stages when illness is typically manifest. Add Health is one of the first large, longitudinal, nationally representative studies to collect biomarker data from a young adult population, when the data may be most informative about pre-disease pathways and cumulative physiological risk. Add Health also has the strengths of extensive interpersonal and physical contextual data, which allow for multilevel analyses of health trajectories from youth into adulthood.

Biomarker Feasibility for Large Scale Studies: Add Health as an Example

Below we describe the theoretical foundation and unique design features of Add Health that facilitate a multilevel systems approach to family process research, followed by our theoretical choices for biological data collection at Add Health, Wave IV.

Theoretical Foundation

The scientific purpose of Add Health is to study developmental and health trajectories across the life course from early adolescence into adulthood using an integrative approach that combines social, behavioral, and biomedical sciences in its research objectives, design, data collection, and analysis. We developed an Integrative Life Course Theoretical Model that specifies three broad conceptual domains of longitudinal and reciprocal influences in trajectories of health and human development: context, behavior, and biology. Here we focus on the biology domain (described below).

Add Health Design

Add Health is an ongoing longitudinal study of a nationally representative sample of more than 20,000 US adolescents who were in grades 7–12 in 1994–95 and have been followed for 15 years through adolescence and the transition to adulthood. In-home interviews occurred in 1995 (Wave I), 1996 (Wave II), 2001–02 (Wave III), and 2008–09 (Wave IV) when Rs were aged 24–32 years. Add Health used a school-based design to obtain direct independent measures of the multiple contexts of young people's lives, including school, peer network, dyadic relationships (friends, romantic and sexual partners), family, neighborhood, and community. Add Health oversampled by ethnicity, physical disability, school (i.e., all enrolled students in 16 schools were selected for in-home interviews), and biological relatedness to increase the diversity of potential research. For example, it sampled > 3,000 pairs of individuals with varying biological relatedness (aka the "genetic sample:" identical/fraternal twins, full & half siblings, cousins, and biologically unrelated youth raised in the same household) to facilitate genetic and environmental research. Thus, Add Health is nationally representative of young people in every race, ethnic, immigrant, geographic, and socioeconomic subgroup now living in all 50 United States. (See Harris 2011 for details.)

Add Health's multilevel and longitudinal data provide unprecedented opportunities to characterize the social, physical, and health environments in the Context Domain of the Integrative Life Course Theoretical Model. Almost 8,000 data elements on the social and physical environment at multiple spatial levels are available across waves (e.g., poverty rates, sexually transmitted infection (STI) prevalence, welfare policies, cigarette taxes, and the proximity and number of physical features such as parks and recreation centers). Add Health also contains extensive longitudinal information for the Behavior Domain measured at multiple levels (i.e., family, school, peer group, relationship dyad, neighborhood, and community), including life histories of sexual and risk behavior, civic engagement, fertility, cohabitation, marriage, and work.

Add Health's design features furthermore enabled measurement in the Biology Domain, including the genetic sample and longitudinal self-reports of physical development, general health, chronic illness, physical activity, mental health, and disability. Add Health has always collected objective measures of health as well, including anthropometrics to identify overweight and obesity, biospecimens to measure STI and HIV status, genetic markers, and more recently, biomarkers of cardiovascular health, metabolic processes, immune function, inflammation, and medication use.

Thus, Add Health provides unique opportunities to study how environments and behaviors are linked to biological and family processes in their influence on health and well-being across the life course. Further, the longitudinal design and diverse sampling strategy enable the examination of bidirectional associations between family characteristics and within-individual change across the life course to support dynamic multilevel system analysis for multiple segments of the US population.

Biology Domain

The choice of biological data in Add Health was driven by scientific knowledge of reasonably prevalent health conditions at a given developmental stage of the Add Health cohort, especially data with implications for future health, the role of specific biological processes in causation, and the ability of specific measures to characterize these processes. Within these scientific criteria, choices were constrained by the feasibility of methods used to obtain valid and reliable physical measurements and biological specimens in a large non-clinical field setting, and measurement strategies and assay techniques used to capture biological phenomena of interest in each life stage. Within these constraints, we chose noninvasive, innovative, cost-efficient, and practical methods for collecting biomarkers appropriate for population research (McDade et al. 2007).

Thus, we measured height and weight in adolescence, during the transition to adulthood, and in adulthood to map the obesity epidemic in Add Health (see Harris 2010; Gordon-Larsen 2010). At Wave III (ages 18–26), the Add Health cohort was at relatively high risk of contracting STIs, so we collected urine and saliva to test for STIs and HIV (see Miller et al. 2004; Morris et al. 2006). We expanded biomarker collection at Wave IV (ages 24–32), when the cohort was settling into adulthood and diverging along pre-disease pathways. We identified obesity, health risk behavior, and stress as the leading health concerns at Wave IV, and collected biological data to provide information about the associated consequences.

Such consequences include hypertension, diabetes, hyperlipidemia, inflammation, immune dysfunction, and cardiovascular disease (Ferraro and Kelley-Moore 2003; McEwen 1998). Therefore, blood pressure and pulse were measured during the Wave IV interview. Metabolic and inflammatory processes associated with obesity also pose significant risks for renovascular, cardiovascular, and cerebrovascular disease (Blake and Ridker 2002; Khaw et al. 2001). Accordingly, we collected capillary whole blood spots from a finger prick onto filter paper, dried, punched and then assayed them for glucose, hemoglobin A1c (HbA1c), lipids (total cholesterol (TC), high-density lipoprotein cholesterol (HDL-C), triglycerides (TG)), and a marker of inflammation, high sensitivity C-reactive protein (hsCRP)—important markers of current health status and future risk for diabetes, kidney disease, peripheral artery disease, heart disease, and stroke (Fagot-Campagna et al. 2001; Nissen et al. 2005). Moreover, the immune system and the hypothalamic-pituitary-adrenal (HPA) axis may be compromised by stressful events, adverse social environments, or health risk behaviors, which in turn can lead to infectious, autoimmune and cardiovascular disease (Herbert and Cohen 1993). We collected saliva to measure a stress hormone, cortisol (pretest only), and assayed dried blood spots for Epstein Barr viral capsid antigen IgG (EBV)—an indirect measure of chronic stress (Glaser et al. 1991). Because genes influence health, behavior, and the moderation of contextual and behavioral effects as they relate to future health, we also collected DNA from all Rs at Wave IV. In combination with the longitudinal contextual and behavioral data, these biological measures inform inter-relationships between biological processes, social, and behavioral trajectories.

Cost of Biomarkers

The financial cost of conducting research involving biomarkers reflects the necessity of many resources. Equipment and supplies are needed for measuring, collecting, storing, shipping, and processing. For example, scales with acceptable levels of accuracy and upper range bounds–and possibly portability, depending on study design–are needed to measure weight. Tubes and syringes are needed for venous blood collection, and materials to keep specimens cold may be needed for packaging and shipment. For some specimens, such as urine or blood, different supplies and/or protocols may be required, depending on the biomarker of interest. Some specimens may be collected relatively easily by Rs themselves (e.g., saliva for buccal cell DNA extraction). However, other types require either specialized field personnel (e.g., nurses or phlebotomists to collect venous blood) or lay interviewers with extensive training (e.g., to conduct skinfold measurements or collect blood spots), which increases costs of measurement. Collection costs also reflect whether specimens are collected at home, at clinic visits, or via mail. Most biomarkers entail laboratory costs to receive, store, and process specimens and to measure the marker of interest. These costs can vary substantially depending on the biomarker and the technology used. In general, the larger the study sample size, the greater the cost. For some biomarkers, such as genetic markers or established hormone assays, the cost typically equals a fixed cost of genotyping times the number of Rs who provided a specimen; for other biomarkers, such as blood pressure or weight, there may be economies of scale if personnel can use the same monitor or scale to obtain measures from multiple Rs in the field. However, costs of other biomarkers can vary substantially, depending partly on whether assays are well-established (e.g., lipids in plasma) or developmental (e.g., lipids in dried blood spots).

Challenges in the Field

Even with extensive research into optimal materials and protocols, Add Health encountered challenges in the field. For example, to avoid keypunch error, digit preference, and data fabrication, we planned to automatically download electronic data from blood pressure monitors. However, electronic downloading was unreliable during the pretest so blood pressures were manually entered in duplicate on laptops. Such trade-offs in performance may be related to the reliance on lower cost equipment required in large scale studies. Clinical monitors from which blood pressure data are reliably downloaded can be ten to twenty times more expensive than the blood pressure monitor used at Add Health Wave IV. Assuming 350 field interviewers, choosing between the two monitor types reflected a difference in total cost of approximately $ 200,000–400,000. Despite such a difference, the monitor used at Wave IV (Model BP3MC1-PC-IB; MicroLife USA, Inc; Dunedin, FL) was approved by the British Hypertension Society and advertised as having an accuracy of 3 mm Hg. Moreover, it performed extremely well in field work (Nguyen et al. 2011).

With non-medical field staff there may also be challenges related to the "ick" factor. Collection of saliva and urine was pre-tested at Wave II of Add Health, but interviewers performed poorly in the field. They were uncomfortable collecting specimens, were afraid to handle them, and may have sabotaged the effort. Given the poor pretest performance, urine collection was abandoned for the main study.

Some specimen collection requires extensive training and adherence to complex protocols. For example, the Wave IV protocol for finger prick whole blood collection required the interviewer to: clean the R's middle or ring finger with an alcohol prep pad, apply a tourniquet to the arm, prick the finger with a lancet, and wipe away the first drop of blood. Next, the interviewer was to drop (ideally) seven blood spots onto a special collection card, ensuring that the finger did not touch the card and that the blood spot saturated the collection circle on the card. Samples were then to be air dried over desiccant for three hours, packaged, and shipped via Federal Express to the lab on the same day as collection. Interviewers had to ensure that they were sending the specimen to the correct lab (Add Health used several labs for different specimens) with the correct R biospecimen ID (different from field interview IDs). The protocol required multiple supplies (plastic gloves, band-aids, gauze, alcohol prep pads, a tourniquet, Lancets, 7-spot collection cards (Whatman 903® *Protein Saver*, Whatman Inc., Piscataway, NJ), Chux pads, and a biohazard container), all of which interviewers carried in the field along with other supplies and a laptop computer.

Seemingly simple saliva collection can also be complicated. For example, although salivary concentrations of most steroid hormones are not affected by saliva flow rate, or mouth collection site, salivary concentrations of some proteins such as alpha-amylase and secretory IgA (SIgA) vary by site (Beltzer et al. 2010; Veerman et al. 1996). Concentrations of some analytes (e.g., SIgA, dehydroepiandrosterone sulfate) vary inversely with saliva flow rates (Kugler et al. 1992; Vining et al. 1983). Further, use of stimulants to increase saliva flow may interfere with assay performance, a concern when Rs self-collect saliva.

Other challenges to specimen integrity over which investigators have little control are exogenous, historical, or cultural events that impact normal field operations. For example, Wave III was in the field during 9/11 (i.e., terrorist attacks on September 11, 2001). This tragic event and its aftermath disrupted the scheduling and keeping of interview appointments (especially on the east coast), but the more enduring effect was severe delays in shipping urine and saliva to labs, resulting in significant biomarker loss. Either specimens were lost in transit, or they arrived at the lab beyond the 48-hour window required for a valid test according to FDA regulations.

Complexity of Protocols for Participants

Stress experiences, and their short- and long-term health implications, were themes of the Wave IV Add Health program project. In the pretest we evaluated a protocol to collect saliva for cortisol measurement. Cortisol is an endogenous corticosteroid that affects multiple physiological systems; it has been implicated in multiple physical and psychological illnesses. Circulating cortisol concentrations can be influenced by stress, but it also has a strong diurnal profile, rising shortly after awakening and then falling throughout the day. Thus, multiple samples are needed, and adherence to the protocol for timing of sample collection is critical. Of particular interest in the context of field collection is the cortisol response to awakening (CRA; a large, rapid increase within a 20–30-minute period after awakening), thought to be a reliable indicator of the acute reactivity of the HPA axis.

Based on available literature, we developed and tested a low burden and affordable protocol to maximize consent and protocol adherence, given that specimen collection would be unsupervised on a day after the in-home interview. We requested that the R collect three specimens on a single day: upon awakening, 30 min later, and just before bed. Rs completed a brief checklist linked to each specimen, noting time of collection, stressful daily events, and recent consumption of food, beverages, or drugs before specimen collection, all needed for proper interpretation of assay results.

Analyses of pretest specimen receipt and protocol adherence were illuminating. Virtually everyone agreed to provide samples, but about a quarter of self-reported collection times were missing for the 76 % of Rs who returned samples, suggesting large amounts of collection time data would be missing in the main field work. Although there was some suggestion that higher incentives ($ 40) might improve specimen return, the increase was cost prohibitive. Given our final sample size of 15,701, applying the 97 % consent and 81 % return rates yields a cost close to $ 494,000 in *incentives alone*. Further, through embedded experiments we determined that collection protocol adherence was poor, thus explaining the awakening sample's intra-class correlation (ICC) of 0.06. Given likely poor data quality, we eliminated saliva collection for cortisol measurement from the main field work (Halpern et al. 2012).

Challenges in the Laboratory

Social scientists may mistakenly assume that, unlike the inherent error and potential bias in survey or observational data, biological measurement is cut and dry, and accuracy is a given. Hormone molecules get "counted," gene allele variations are clear, and all procedures are standardized and scrupulously followed by infallible laboratory technicians. Alas, this is simply not true. Below we illustrate some of the potential pitfalls that can occur.

Technology Change, Time, and Processing Delays

Equipment, techniques, reagents, etc. evolve. Whereas DNA and hormones were once measured only in blood from venous draws, many are now obtained via buccal cells and saliva. However, data derived from new technology must be appropriately scrutinized. For example, when salivary radio-immunoassays (RIAs) for steroids were introduced, many assumed these early assays had the same accuracy and reliability as RIAs using blood samples. In practice, this was not necessarily true. In one experiment, correlations of salivary testosterone (T) measurement from the same specimens across labs ranged from 0.05 to 0.48 (Halpern and Udry 1992). Similarly, correlations of T values between plasma and saliva samples collected from the same person at the same point in time ranged across labs from −0.009 to 0.86; even within a lab the correlation was only 0.52 (Halpern and Udry 1992). Perhaps such differences simply introduce noise and bias biological associations with behavior toward the null; however, this is not necessarily true. To illustrate, correlations between male adolescents' reports of "frequency of thinking about sex" and salivary T ranged from an insignificant 0.12 to a statistically significant 0.40, depending on lab measurement used (Halpern and Udry 1992).

There are other challenges related to state-of-the-art laboratory methods. For example, we assayed lipids in dried whole blood spots at Wave IV. Originally, the lipids were colorimetrically assayed using procedures that measure change in color (optical density) reflective of increases in plasma lipid concentrations. However, the colorimetric assays were replaced with fluorimetric alternatives during specimen collection and assay. The anticipated advantage (e.g., validity) of the fluorimetric assays—which involve ultraviolet excitation of and spectroscopic measurement of light emitted from fluorochromes—led to their adoption. That said, subsequent examination of the temporal variation, short-term reliability, inter-convertibility, and internal/external validity of the two assays did not consistently substantiate the purported advantage of the fluorimetric assays.[1]

Factors that delay assaying or genotyping samples (e.g., overwhelmingly large numbers) make consistency of laboratory procedures and uniformity of supplies/reagents an issue. Indeed, assay performance may vary over time and among supply/reagent lots while measurement tools such as genotyping platforms may be retired. These possibilities can be difficult to avoid if field work is prolonged, or if a lab has inadequate automation to accommodate incoming sample volume.

Art and Ambiguity

Genotyping is another domain challenged by the state-of-the-art of measurement. There are issues related to the genotyping chemistry and software used to differentiate true alleles from various sources of background error. (An allele is one of two or more forms of a gene or genetic locus.) Variation in allele calls, and therefore the labeling of genotypes, is a good example. Calling is based on a pattern of peaks, points, or bands on a computer generated "image" of an electropherogram or gel. Although most laboratories use software programs for sizing (repeat polymorphisms) and clustering (single nucleotide polymorphisms or SNPs), some human judgment is still involved. To assist judgment software packages include detailed sections on manual editing of automated allele calls. For sizing polymorphisms, the two most common problems are "stutter bands" (strong peaks that are one or more repeat units smaller than the actual allele) and allele dropout (large alleles poorly amplified and missed). The editing process involves two independent reviewers of peak profiles who may offer discordant calls that must be reconciled. If not reconciled, the genotype will be judged to be "missing" (Hill et al. 2004).

Virtually all SNP genotyping, from a single Taqman assay to multi-million-SNP chips, is based on assigning genotypes to clusters of points visualized on a Cartesian plane. Ideally, one homozygote forms a tight cluster on the X axis, the opposite homozygote clusters on the Y axis, and the heterozygotes cluster neatly in between. Manual editing is required when the software cannot define clusters accurately or cannot assign genotypes to points not clearly in one of the three clusters. For relatively small arrays this is feasible, but for large arrays it is impossible. Multiple software applications have been developed that claim to call genotypes more accurately than the manufacturer-supplied software (e.g., Browning and Yu 2009), but the implication of these improvements is clear: The genotypes for any set of samples genotyped on very large arrays will vary depending on the software package used to define them.

[1] However, a separate inter-conversion process for HbA1c (necessitated by a post-pretest lab closing) was successful (see Table 8.1).

Table 8.1 Reliability of Add Health Wave IV Biomarkers Based on IIV Study (Details of Add Health Wave IV equipment and measurement protocols are available at http://www.cpc.unc.edu/projects/addhealth/data/guides)

Type	Measure	ICC (95 % CI)[a]
Anthropometric	Weight	1.00 (1.00–1.00)
	Height	0.98 (0.98–0.99)
	BMI	0.99 (0.99–1.00)
	Waist	0.98 (0.97–0.99)
Cardiovascular	SBP	0.81 (0.74–0.88)
	DBP	0.68 (0.57–0.79)
	PR	0.47 (0.31–0.63)
Metabolic	HbA1C	0.97 (0.96–0.98)
	Glucose (non-fasting)	0.39 (0.21–0.58)
	TG	0.71 (0.60–0.81)
	TC	0.40 (0.22–0.58)
	HDL-C	0.31 (0.12–0.51)
Inflammatory	hsCRP	0.70 (0.59–0.81)
Immune	EBV	0.97 (0.96–0.98)

[a] *ICC (95 % CI)* intra-class correlation coefficient, 95 % confidence interval

Challenges of Scale

In a large study with concentrated data collection, questions of scale become critical. Factors that delay processing, assaying, or genotyping samples (e.g., overwhelmingly large numbers, inadequate automation) make consistency of laboratory procedures and uniformity of supplies/reagents an issue. Assay performance may vary over time and among supply/reagent lots; measurement tools such as genotyping platforms may be retired. An inadequately staffed or automated lab will be hard pressed to process the *hundreds* of specimens that will arrive *daily*. Processing issues may result in labeling/storage errors, contamination of samples, errors in data entry, or exceeding the time window for storage outside of a −70 C freezer. Further, depending on the size and number of specimen aliquots, simple storage space may be an issue, both for the lab and for the investigator, if the specimens are to be archived.

Quality Control

One unique aspect of the Add Health Wave IV design was to include an intra-individual variation (IIV) study. Inclusion of an IIV study involved repeated collection of biomarkers on the same individuals over a short time interval to estimate their reliability. Approximately 100 IIV Rs were interviewed twice, one to two weeks apart. The first visit included a full interview and full set of biomarkers. The second visit included an abbreviated interview, mainly capturing information needed for biomarker interpretation, and a full set of biomarkers. Labs and technicians were blinded to IIV specimens. We computed ICCs as measures of reliability, then used them to monitor biomarker data quality and correct for measurement error.

Table 8.1 shows reliability information derived from the IIV study. The ICCs of course vary depending on the stability and susceptibility of a given biomarker to behavioral and/or contextual change. Thus, height, weight, and EBV, for example, have excellent test-retest reliability, while more labile measures such as pulse rate were lower. Comparisons with external standards were conducted to monitor biomarker validity (see http://www.cpc.unc.edu/projects/addhealth/data/guides/WaveIV).

Biomarker Consent and Compliance

Investigators may wonder what level of biomarker consent to expect, and whether requests for specimen collection will affect consent rates for study participation. In Add Health, the longitudinal design has reinforced consistent rapport with Rs, who trust the rigorous security system we maintain to ensure our original pledge of confidentiality to them. An average of 99 % of Rs agreed to height and weight at all waves (including waist circumference at Wave IV). Compliance for Wave IV blood pressure readings was also 99 %. Consent rates at Wave III were also high: 92 % provided urine for STI testing; 95 % provided saliva for HIV testing; and 83 % provided buccal cell saliva for DNA testing. The lower compliance for DNA may be due to not providing a separate monetary incentive for DNA at Wave III as was done for urine and saliva.

At Wave IV we obtained signed consent separately for the collection of salivary buccal cell DNA, blood spots, and salivary cortisol (pre-test only) at the beginning of the interview when we obtained signed consent for the Wave IV survey administration. This allowed Rs to change their minds about biomarker consent at the end of the survey and before biomarker collection. (A few Rs did, in both directions). We also used a two-tiered consent process for (1) currently planned Program Project research and (2) future research "related to long term health." The latter provided a biospecimen archive for future testing. Consent to biospecimen collection was uniformly high: 96 % for DNA and 95 % for blood spots as part of the planned program project research. Consent to archival for future analysis also was high: 78 % for DNA and 80 % for blood spots. Black and Asian Rs were somewhat less likely to consent to biospecimen archival than Hispanic and white Rs.

Biospecimen Archive

A significant scientific advantage of collecting biospecimens in a large-scale survey is the potential for assembling and maintaining an archive or "bio-repository." The Add Health Biospecimen Archive includes urine (Wave III), DNA (Wave III and Wave IV), and dried blood spots (Wave IV). Planning for an archive requires additional tasks and costs including (1) a separate consent process for archival, (2) budgeting for long-term storage and additional shipping costs and logistics for returning samples from laboratories, (3) greater field and training demands to procure enough specimen to archive, (4) development of dissemination policies for sharing specimens, and (5) plans for the archive if/when the study ends. However, the scientific benefits of maintaining an archive far outweigh the costs. The archive allows study investigators and outside investigators, through an ancillary study mechanism, to capitalize on rapidly changing technology to produce additional biomarker information that can be linked with the existing survey, geographic, and biological longitudinal data.

In the last 10 years new methods have been rapidly developed to analyze saliva and dried blood spot samples for an increasing array of biomarkers (McDade et al. 2007) and to explore a widening range of specimens collected in field settings (e.g., hair, fingernails, vaginal swabs). During the last decade enormous advances in genotyping technology, including chip arrays that accommodate in excess of 2 million genetic markers, have been made. In pace with technological change has been a parallel decline in cost and an explosion of new research and knowledge. As both technology and scientific discovery advance, an archive offers opportunities to test new hypotheses in both cross-sectional and longitudinal studies.

Dissemination Challenges

A biospecimen archive raises numerous dissemination issues including data sensitivity (e.g., diabetes markers, STI results, genetic data), deductive disclosure risks when sensitive biomarker data are merged with survey and geographic data, security requirements for access, analysis, and publication of the data with which users must comply, and determining who should have access and under what conditions (e.g., what forms of the data will be released, what merged files are permissible, what level of review and oversight of user access is needed)[2]. Biomarker data also increase knowledge demands (and therefore demands for training and experience) on users in terms of interdisciplinary understanding, research design, and analytic methods that enable users to develop meaningful research questions, appropriately use and model the data, and properly interpret their findings. Wide access to biomarker data will undoubtedly increase the possibility that inexperienced users will misuse it or misinterpret results.

Fortunately, there is growing research and information technology expertise on how to securely release sensitive data to minimize disclosure risks. Add Health pioneered an innovative dissemination policy, later adapted by other studies, to release data using a four-tiered access plan whereby the security requirements for access and use become stricter as the data requested represent greater disclosure risks or require more monitoring of use (e.g., genetic data versus blood pressure or glucose concentrations). Providing hands-on workshops or didactic sessions on biomarker data (e.g., as done at the Add Health Users Conferences) can improve the proper use of diverse types of biomarkers.

---

[2] Space limitations preclude elaboration here but we also want to alert readers to the complex issues surrounding the obligation to report (especially urgent and emergent) laboratory values with established clinical utility to respondents or their designees, and the simultaneous obligation not to harm while doing so.

Recommendations for Best Practices

Our experiences suggest the following practices to optimize biomarker data quality:

- Read available literature and consult with experts to evaluate the biomarkers, and their collection and processing protocols that are best suited for your research aims and design.
- Use standardized collection protocols and pretest them exactly as they are to be implemented.
- Collect more specimen than you need and avoid storing or shipping all specimens together. Assays may need to be repeated and natural disasters can wipe out an entire archive.
- Implement uniform training, certification, and retraining of staff; use identical supply/reagent/assay types and sources; and conduct regular equipment testing and calibration.
- Never assume that your lab has caught all the problems or issues, especially when assays or techniques are relatively new, or are new to the lab. Never assume that field staff and/or Rs are following the protocol correctly. Introduce your own real-time, masked, external quality control and assurance procedures to assess reliability and validity. A measuring tape placed in the wrong location on the body, even just slightly off, can change measures dramatically.
- Be clear about how long specimens are to be retained, and under what storage conditions.

Illustrative Insights from Biomarkers

As large scale studies begin to collect biomarkers, research opportunities to integrate new biological data with social, behavioral, and environmental data in a multi-system model are being realized. Here we provide a few examples of this integrative research. MIDUS is an ongoing longitudinal aging study of adults aged 25–74 years in 1995 designed to investigate the role of behavioral, psychological, and social factors in accounting for age-related variations in health and well-being in a national sample (http://midus.wisc.edu/index.php). Integrating biological data used to measure metabolic syndrome in adulthood (i.e., waist circumference, blood pressure, lipids, blood glucose) with retrospective survey data on socioeconomic status and parents' behavior in childhood, Miller et al. (2011) report a buffering effect of maternal nurturance in the influence of childhood poverty on metabolic syndrome such that high levels of maternal nurturance offset the metabolic consequences of childhood disadvantage.

Another recent study examined how transitions to fatherhood impact testosterone (T) levels using longitudinal survey data and biological samples collected in a large sample of men participating in the Cebu Longitudinal Health and Nutrition Survey (CLHNS), a representative 1-year birth cohort study begun in the Philippines in 1983

Table 8.2 Prevalence estimates of selected health conditions using survey, biomarker and pharmacologic data, young adults ages 24–32 (2008–09) (national longitudinal study of adolescent health (Wave IV))

Health Condition	Percentage (%)
Hypertension (*N = 14,252*)	
Use medication	3.4
Self-reported	11.1
Use medication or self-reported	11.8
Use medication, self-reported, SBP $\geq$ 160 or DBP $\geq$ 100[a]	14.0
Use medication, self-reported, SBP $\geq$ 140 or DBP $\geq$ 90[a]	26.1
Diabetes (*N = 12,224*)	
Use medication	1.3
Self-reported	2.5
Use medication or self-reported	2.8
Use medication, self-reported, or glucose $\geq$ 200[b]	3.4
Use medication, self-reported, glucose $\geq$ 200, or HbA1c $\geq$ 6.5[b]	5.5

[a] Stage 2 hypertension is classified as SBP $\geq$ 160 or DBP $\geq$ 100; Stage 1 hypertension is classified as SBP $\geq$ 140 or DBP $\geq$ 90 ("The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure (JNC 7)." 2003. *Hypertension* 42: 1206)
[b] Random (non-fasting) glucose values $\geq$ 200 mg/dL and HbA1c values $\geq$ 6.5 % are cut-offs for classification of diabetes (American Diabetes Association. 2007. "Diagnosis and Classification of Diabetes Mellitus." *Diabetes Care* 30(S1): S 42–47)

(http://www.cpc.unc.edu/projects/cebu). Salivary T was assessed when the cohort was aged 21.5 (in 2005) and again at age 26 (in 2009). Men with the highest levels of T in 2005 were more likely to become committed partners and fathers by 2009, but those who did become fathers showed steeper drops in T compared to their single, childless counterparts (Gettler et al. 2011). Testosterone levels were lowest in men who spent the greatest amount of time caring for their children. These findings demonstrate the bi-directional relationships between family and biological processes.

As mentioned earlier, self-reports tend to underestimate the population prevalence of health conditions, especially among young people who are in a healthy life stage and for conditions that are asymptomatic. Table 8.2 shows prevalence estimates of hypertension and diabetes at Wave IV, when the Add Health cohort was in their mid-20s and early 30s. By combining self-reported, biological, and medication data, we obtain more sensitive estimates. In this young adult population, 11.1 % self-report hypertension. Combined with antihypertensive medication use, prevalence rises slightly to 11.8 %. Bringing in blood pressures (BPs) and using conventional thresholds to define hypertension according to Joint National Committee (JNC) 7 guidelines (Chobanian et al. 2003), the prevalence of stage 2 hypertension rises to 14.0 %; and stage 1 hypertension rises to more than 25 %. We see similar measurement gains in prevalence of diabetes as defined by the American Diabetes Association (ADA 2012). These data illustrate the insensitivity of self-reported health data, and related measurement error present in modeling these disease outcomes, especially among seemingly healthy young populations.

The integration of genetic and social data to understand how the environment moderates gene expression in behavioral outcomes has captured the attention of the social science research community. Guo et al. (2010) examined how the dopamine transporter gene (*DAT1*) interacts with age (or life course stage) in relation to risk behavior (delinquency, number of sex partners, substance use, and seatbelt use) from adolescence into young adulthood, using data from Waves I, II and III of Add Health. They reported a protective effect of the 9R/9R genotype in the VNTR of *DAT1* on risky behavior. However, this protective effect varied according to age/life course stage, such that genetic protection is evident when the risk behavior is illegal (e.g., alcohol use and smoking in adolescence) but vanishes when the behaviors are legal or more socially tolerated (e.g., alcohol use and smoking in adulthood). This important research demonstrates how legal, as well as social, contexts can moderate genetic associations for diverse behaviors.

A final example relates to STIs. Bruckner and Bearman (2005) examined the effectiveness of adolescent virginity pledges (i.e., a pledge to remain a virgin until married) in reducing STI rates among young adults aged 18–24 years in Add Health. Urine specimens collected at Wave III were tested for human papilloma virus, chlamydia, gonorrhea, and trichomoniasis; positive results indicated infection. Pledgers were consistently less likely to be exposed to risk factors across a wide range of indicators, but their STI prevalence did not differ from non-pledgers. Bruckner and Bearman (2005) hypothesized that pledgers were less likely to use condoms at sexual debut and less likely to be tested, and therefore diagnosed, with STIs. They concluded that any health advantages accruing from pledging do not appear to stem STI acquisition among young adults.

In an unpublished response to this research, Rector and Johnson (2005) re-examined the linkage between adolescent virginity pledging and STIs among young adults using self-reported data on diagnosis or symptoms at Wave III. Their results are opposite to Bruckner and Bearman, finding that virginity pledging predicts lower STI prevalence among young adults when STIs are measured by self-report. However, as illustrated above, self-reports fail to capture many health conditions that are better indexed by objective biological measures. These two sets of findings are a compelling example of how different public policy implications can be depending on the measures used.

Now, Tell me Again Why I Should Consider Biomarkers?

After learning about the expense, challenges, and complications inherent in biomarker collection on a large scale, one might ask again, “Why use biomarkers?” Despite the challenges, there are many advantages to using biomarkers as measures of health and family process. First, biomarker data provide information about health conditions that may be unknown to the individual. Being unaware of potential health problems is especially likely in young adulthood when, being without disease symptoms, young people assume they are healthy and are unlikely to have regular medical check-ups. A second advantage is the opportunity to archive biospecimens for future analysis to capitalize on advances in technology, new research knowledge, and potentially declining costs for testing. This is especially relevant for DNA archives where genotyping techniques are changing at escalating speed while costs decline.

Third, repeated biomarker assessments in longitudinal studies allow for analysis of change, including disease onset and progression, and the ability to map predisease pathways. This is extremely valuable in young populations before disease is manifest, because identifying the precursors to disease will lead directly to policy and program interventions to improve health and lower disease prevalence. Fourth, biomarkers reflect different time metrics and therefore offer different types of insight into health status and biological processes. For example, some markers measure current status (e.g., STI, blood glucose), while others measure cumulative health (e.g., HIV status, diabetes) or contextually dependent change (e.g., cortisol). A fifth advantage of using biomarkers is that biomarkers are not influenced by recall bias or by individual characteristics that tend to bias self-reports of health.

As J. Richard Udry advocated throughout his long career, biosocial theory and research have already made, and will continue to make, important contributions to our knowledge of sociological problems particularly issues related to family functioning, and the future holds even greater opportunities for biosocial family research. For example, Add Health is expanding its family system design with intergenerational investigation of health and development across three generations: the parents of Add Health Rs (G1); Add Health Rs (G2); and the children of Add Health Rs (G3). The long-term goal is to collect parallel social, environmental, behavioral, biological, and genetic data on three generations to enable unprecedented multilevel dynamic systems research on the intergenerational linkages in health and behavior in family systems. We encourage other family researchers who have not considered the inclusion of biomarkers in their research to broaden their scientific reach as well.

Authors' Note We dedicate this work to J. Richard Udry, a pioneer in biosocial research, the original Director of Add Health, and a mentor to Halpern, Harris, and many others who had the good fortune to work with him.

References

American Diabetes Association. (2012). Diagnosis and classification of diabetes mellitus. *Diabetes Care, 35,* 64–71.

Beltzer, E. K., Fortunato, C. K., Guaderrama, M. M., Peckins, M. K., Garramone, B. M., & Granger, D. A. (2010). Salivary flow and alpha-amylase: Collection technique, duration, and oral fluid type. *Physiology and Behavior, 101*(2), 289–296.

Blake, G. J., & Ridker, P. M. (2002). Inflammatory biomarkers and cardiovascular risk prediction. *Journal of Internal Medicine, 252*(4), 283–294.

Browning, B. L., & Yu, Z. (2009). Simultaneous genotype calling and haplotype phasing improves genotype accuracy and reduces false-positive associations for genome-wide association studies. *American Journal of Human Genetics, 85*(6), 847–861.

Brückner, H., & Bearman, P. (2005). After the promise: The STD consequences of adolescent virginity pledges. *Journal of Adolescent Health, 36*(4), 271–278.

Chobanian, A. V., Bakris, G. L., Black, H. R., Cushman, W. C., Green, L. A., Izzo, J. L., National High Blood Pressure Education Program Coordinating Committee (2003). Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Hypertension, 42*(6), 1206–1252.

Fagot-Campagna, A., Saaddine, J. B., Flegal, K. M., & Beckles, G. L. (2001). Diabetes, impaired fasting glucose, and elevated HbA1c in U.S. adolescents: The third national health and nutrition examination survey. *Diabetes Care, 24*(5), 834–837.

Ferraro, K. F., & Kelley-Moore, J. A. (2003). Cumulative disadvantage and health: Long-term consequences of obesity? *American Sociological Review, 68*(5), 707–729.

Finch, C. E., Vaupel, J. W., & Kinsella, K. (Eds.), Committee on Population, National Research Council. (2001). *Cells and surveys: Should biological measures be included in social science research?* Washington: National Academies Press.

Gettler, L. T., McDade, T. W., Feranil, A. B., & Kuzawa, C. W. (2011). Longitudinal evidence that fatherhood decreases testosterone in human males. *Proceedings of the National Academy of Sciences, 108*(39), 16194–16199.

Glaser, R., Pearson, G. R., Jones, J. F., Hillhouse, J., Kennedy, S., Mao, H. Y., & Kiecolt-Glaser, J. K. (1991). Stress-related activation of Epstein-Barr virus. *Brain, Behavior, and Immunity, 5*(2), 219–232.

Gordon-Larsen, P., The, N. S., & Adair, L. S. (2010). Longitudinal trends in obesity in the United States from adolescence to the third decade of life. *Obesity, 18*(9), 1801–1804.

Gottlieb, G., & Halpern, C. T. (2002). A relational view of causality in normal and abnormal development. *Development and Psychopathology, 14*(3), 421–435.

Guo, G., Cai, T., Guo, R., Wang, H., & Harris, K. M. (2010). The Dopamine Transporter Gene, a Spectrum of Most Common Risky Behaviors, and the Legal Status of the Behaviors. *PLoS ONE, 5*(2), e9352. doi:10.1371/journal.pone.0009352.

Halpern, C. T., & Udry, J. R. (1992). Variation in adolescent hormone measures and implications for behavioral research. *Journal of Research on Adolescence, 2*(2), 103–122.

Halpern, C. T., Whitsel, E. A., Wagner, B., & Harris, K. M. (2012). Challenges of measuring diurnal cortisol concentrations in a large population-based field study. *Psychoneuroendocrinology, 37*(4), 499–508.

Harris, K. M. (2010). An integrative approach to health. *Demography, 47*(1), 1–22. Accessed 17 August 2012.

Harris, K. M. (2011). Design features of add health. http://www.cpc.unc.edu/projects/addhealth/data/guides/design%20paper%20WI-IV.pdf.

Herbert, T. B., & Cohen, S. (1993). Stress and immunity in humans: A meta-analytic review. *Psychosomatic Medicine, 55*(4), 364–379.

Hill, S. Y., Shen, S., Zezza, N., Hoffman, E. K., Perlin, M., & Allan, W. (2004). A genome wide search for alcoholism susceptibility genes. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 128B*(1), 102–113.

Kehoe, R., Wu, S. Y., Leske, M. C., & Chylack L. T. Jr., (1994). Comparing self-reported and physician-reported medical history. *American Journal of Epidemiology, 139*(8), 813–818.

Khaw, K. T., Wareham, N., Luben, R., Bingham, S., Oakes, S., Welch, A., & Day, N. (2001). Glycated hemoglobin, diabetes, and mortality in men in Norfolk cohort of European Prospective Investigation of Cancer and Nutrition (EPIC-Norfolk). *BMJ, 322,* 15–28.

Kugler, J., Hess, M., & Haake, D. (1992). Secretion of salivary immunoglobulin A in relation to age, saliva flow, mood states, secretion of albumin, cortisol, and catecholamines in saliva. *Journal of Clinical Immunology, 12*(1), 45–49.

McDade, T. W., Williams, S., & Snodgrass, J. J. (2007). What a drop can do: Dried blood spots as a minimally invasive method for integrating biomarkers into population-based research. *Demography*, *44*(4), 899–925. doi:10.1353/dem.2007.0038.

McEwen, B. S. (1998). Stress, adaptation and disease. Allostasis and allostatic load. *Annals of the New York Academy of Sciences, 840,* 33–44.

Miller, W. C., Ford, C. A., Morris, M., Handcock, M. S., Schmitz, J. L., Hobbs, M. M., Udry, J. R. (2004). The prevalence of chlamydial and gonococcal infections among young adults in the United States. *The Journal of the American Medical Association, 291*(18), 2229–2236.

Miller, G. E., Lachman, M. E., Chen, E., Gruenewald, T. L., Karlamangla, A. S., & Seeman, T. E. (2011). Pathways to resilience: Maternal nurturance as a buffer against the effects of childhood poverty on metabolic syndrome at midlife. *Psychological Science, 22*(12), 1591–1599. doi:10.1177/0956797611419170.

Morris, M., Handcock, M. S., Miller, W. C., Ford, C. A., Schmitz, J. L., Hobbs, M. M., & Udry, J. R. (2006). Prevalence of HIV infection among young adults in the United States: Results from the Add Health study. *The American Journal of Public Health, 96*(6), 1091–1097.

Nissen, S. E., Tuzcu, E. M., Schoenhagen, P., Crowe, T., Sasiela, W. J., Tsai, J., Ganz, P. (2005). Statin therapy, LDL cholesterol, C-reactive protein, and coronary artery disease. *New England Journal of Medicine, 352*(1), 29–38.

Nguyen, Q. C., Tabor, J. W., Entzel, P. P., Lau, Y., Suchindran, C., Hussey, J. M., Whitsel, E. A., et al. (2011). Discordance in national estimates of hypertension among young adults. *Epidemiology, 22*(4), 532–541.

Rector, R., & Johnson, K. A. (2005). *Adolescent virginity pledges, condom use and sexually transmitted diseases among young adults*. Paper presented at the National Welfare Research and Evaluation Conference, Washington, DC.

Udry, J. R. (1988). Biological predispositions and social control in adolescent sexual behavior. *American Sociological Review, 53*(5), 709–722.

Veerman, E. C., van den Keybus, P. A., Vissink, A., & Nieuw Amerongen, A. V. (1996). Human glandular salivas: Their separate collection and analysis. *European Journal of Oral Sciences, 104*(4), 346–352.

Vining, R. F., McGinley, R. A., & Symons, R. G. (1983). Hormones in saliva: Mode of entry and consequent implications for clinical interpretation. *Clinical Chemistry, 29*(10), 1752–1756.

Weinstein, M., Vaupel, J. W., & Wachter, K. W. (Eds.), Committee on Advances in Collecting and Utilizing Biological Indicators and Genetic Information in Social Science Surveys, National Research Council. (2008). *Biosocial surveys*. Washington: National Academies Press.

Chapter 9
Ecological Momentary Assessment (EMA) in Family Research

Joshua M. Smyth and Kristin E. Heron

Ecological Momentary Assessment (EMA) in Family Research

Approaches to collecting naturalistic data in daily life have evolved to meet research needs in a variety of disciplines. Family researchers are often interested in dynamic, complex processes that occur both within and between individuals, are present in daily life and regular activities, and often unfold over periods of time. One approach is to repeatedly, and often intensively, collect information on people as they live their normal daily lives. The idea of periodically signaling people to report on experiences as they go about their daily lives was known early on as the Experience Sampling Method (ESM) (Csikszentmihalyi and Larson 1987). Somewhat more recently, this approach broadened to include reports not only of internal states, but also of behaviors, social and physical context, and other variables, and became known as Ecological Momentary Assessment (EMA). EMA, ESM, and other approaches constitute a range of methodological tools that enable researchers to collect large amounts of information in people's natural environments. As it reflects both the approach and the nomenclature we are most familiar with, we will use the terminology EMA for the remainder of the chapter.

Information can be collected from people in their daily lives using a variety of strategies, including collecting information in response to specific events, at predetermined times, or at seemingly random times. Before mobile electronic technology was widely available, participants were given pagers that were programmed to "beep" at random times throughout the day, and participants would complete assessments in paper diaries they were required to carry with them (Scollon et al. 2003; Stone et al.

J. M. Smyth (✉)
Departments of Biobehavioral Health & Medicine,
The Pennsylvania State University, University Park, PA, USA
e-mail: jms1187@psu.edu

K. E. Heron
Survey Research Center, The Pennsylvania State University,
University Park, PA, USA
e-mail: keh16@psu.edu

S. M. McHale et al. (eds.), *Emerging Methods in Family Research,*
National Symposium on Family Issues 4, DOI 10.1007/978-3-319-01562-0_9,

2007). As the sophistication and affordability of mobile technology has improved, devices such as palmtop computers (i.e., personal digital assistants (PDAs)), mobile phones, and smartphones are commonly used for EMA studies. In the following sections, we will discuss EMA strategies and the advantages and challenges of using mobile devices as data collection tools with an emphasis on how these issues may be relevant to family research.

Rationale for Using EMA

There are many reasons researchers may desire intensive longitudinal data for their research. In general, the rationale for using EMA rests on three central benefits of these methodologies: (1) reducing retrospective recall/memory biases, (2) facilitating ecological validity, and (3) the capacity to examine within person processes and temporal dynamics.

Reducing Retrospective Memory Biases

Most family researchers use some self-report data (from questionnaires or interviews), often in conjunction with other data, a point we return to below. Self-report data are usually collected via retrospective assessments, asking participants to summarize experiences over some time period (weeks, months, years). Extensive evidence indicates that such information recall is affected by heuristics used in memory search and reconstruction (Scollon et al. 2003; Smyth and Stone 2003), which can systematically bias participant responses. For instance, people are more likely to recall salient and intense/peak experiences, and tend to remember past events in accordance with what they currently know about the event or behavior (Hufford 2007; Smyth and Stone 2003; Stone and Shiffman 1994). Because EMA typically asks participants to report on their current state or very recent experiences (minutes or hours), retrospective recall biases are often greatly reduced.

Facilitating Ecological Validity

EMA is typically implemented in participants' natural settings, or daily life, thus greatly increasing the likelihood for ecological validity and generalizability. Research and clinical settings (e.g., laboratory, hospital, etc.) are often quite artificial environments in the sense that they do not reflect either the typical or full range of individuals' lives. It can be difficult or even impossible to assess important behaviors challenging to observe or create in the clinic/laboratory. Furthermore, it is unclear whether relationships identified in research or clinical settings are necessarily similar to those in individuals' everyday lives (Smyth and Stone 2003). For example, there is evidence that some people are classified as having hypertension when blood pressure is taken in a medical setting, but their blood pressure in daily life (i.e., outside of the clinic) is normal. This suggests that the assessment conducted in the clinic is not representative of the individual's blood pressure in their normal environment (so-called "white-coat hypertension"; Pickering and Freedman 1991). Thus, EMA approaches allow bypassing the assumption that what is observed in the laboratory is generalizable to daily life. (It should be noted, however, that there are cases where the very presence of such "disagreement" between lab/clinic and daily life is clinically and/or theoretically informative!)

Examining Within-Person Relationships and Temporal Dynamics

Intensive assessment strategies that collect multiple assessments (across days or throughout a day) can allow researchers to explore within person relationships and temporal dynamics. For example, many intra-personal (e.g., mood, stress) and inter-personal (e.g., supportive or hostile exchanges) processes are quite dynamic, changing rapidly over short time frames (e.g., minutes, hours). The intensive data capture provided by EMA allows for more than examining person-averages of such processes, but rather opens up the capability to model the temporal dynamics of these processes (and the many potential influences researchers may be interested in examining). For instance, EMA methods have been used to identify the affective antecedents and consequences of eating disorder behaviors (binge eating and purging) in young women with bulimia nervosa (Smyth et al. 2007) and in children aged 8–13 with binge eating (Hilbert et al. 2008). In both studies, participants were prompted to completed EMA assessments on palmtop computers or cellular phones at semi-random times throughout the day and also after eating episodes. Results showed that stress and negative affect (Smyth et al. 2007) and negative thoughts about food (e.g., dieting, food cravings) and about one's body (e.g., body dissatisfaction; Hilbert et al. 2008) were reported at higher rates prior to and following binge eating episodes than at randomly prompted times. Similar approaches have been used to study the real-world "triggers" of headache onset in children aged 8–17 with chronic headaches. Participants completed electronic diaries three times daily for 2 weeks. Results showed that increases in reported stress immediately preceded the onset of new headache episodes (Connelly and Bickel 2011). As these examples demonstrate, an important benefit of using methods that collect multiple daily measurements is that complex and nuanced questions about dynamic associations and processes that occur over time can be addressed. Moreover, if the data capture has a sufficient number of observations, it becomes possible to model change and relationships between variables in a dynamic fashion (e.g., allowing estimates of non-linear processes, recursive and other feedback effects, etc.).

Although EMA can be used to evaluate dynamic changes over relatively short time frames, daily assessment methods can also be coupled with more traditional longitudinal assessment designs to assess processes over longer developmental intervals. Measurement burst designs provide for intensive "bursts" of assessments collected over short periods of time (days, weeks) to then be repeated for the same individuals over longer time intervals (e.g., months, years; see Sliwinski 2008). For instance, Sliwinski et al. (2009) used measurement-burst diary studies to assess emotional reactivity to daily stress across the lifespan. In these studies a burst of daily assessments of mood and stress were collected and then repeated after a 10-year interval (study 1) or every 6 months for 2 years (study 2). By using these two studies, both relatively long (10 year) and short (2 year) developmental periods could be examined. Results showed that, although there were stable aspects to individuals' stress reactivity, it increased over time during the adult lifespan. Moreover, individuals also showed greater *daily* stress reactivity during times in their life when they were experiencing greater *global* stress, demonstrating how processes at one level can influence the relationship between variables at another level.

These intensive daily and within day data capture techniques thus allow for the within-person association between variables to be well-characterized over short periods. Through combining these data capture methods with other approaches, processes can be characterized over much longer developmental periods. The collection of intensive longitudinal data can represent an important improvement over the use of cross-sectional data (where of course no temporal sequencing can be demonstrated), and—at least in some methodologies, such as EMA—can provide data capture of sufficient density to conceptually and statistically model the level of temporal analysis desired (be that days, hours, or minutes). Thus, a range of research questions can be examined in ways not possible with other data capture approaches (such as understanding response and recovery times, effect durations, recursive dynamics, etc.).

Methodological Consideration with EMA

The following sections provide an overview of several of the issues to be considered when designing an EMA study. Specifically, we describe the possible "targets" of assessment (i.e., who is being assessed), various assessment protocols that can be used, and the types of constructs that can be assessed by EMA.

Assessment Targets: Who is Being Assessed Using EMA?

EMA methods can be used to collect real-time data about individuals from a variety of perspectives. Self-report data about ones' self, other people, and environmental or contextual factors can all be obtained and this may be done by multiple people/informants. In family research, there are several examples of investigators using

EMA to collect intensive, real-world data from both children and their parents to capture experiences from multiple perspectives. For instance, in a study designed to elucidate the challenges of everyday family life in a sample of children with ADHD, mothers and children independently rated location, current activity, social context, and the child's behaviors (e.g., restless, impatient) and moods (e.g., angry); mothers also reported their own moods (Whalen et al. 2006). This study provides an example of how EMA methods can be used to simultaneously collect reports from multiple family members, as well as examples of the variety of "types" of information that can be collected; the mothers in this study provided self-report data about themselves (mood), others (child's behavior), and environmental variables (social context).

Assessment Timing: When Can EMA Reports Be Collected?

There are a range of assessment timing strategies for EMA. Most researchers use one or more of three approaches: (1) completing surveys at predetermined times (e.g., every 4 h; *interval-contingent*), (2) signaling participants to complete EMA reports (e.g., an audible tone or SMS message; *signal-contingent*), and (3) participant initiated assessment in response to a specific event (e.g., a parent-child interaction; *event-contingent*). Importantly, such strategies can be used alone, but are often combined with each other (and/or with other data collection methods; e.g., observation, interview, etc.). Each approach is described below.

Interval-contingent measurement is used when researchers are interested in specific time periods, which may or may not be linked to events. For example, a researcher could have a new parent complete an EMA survey in the morning after waking up (e.g., with questions about the upcoming day or previous night's sleep), or immediately before going to bed (e.g., assessing global impressions of experiences that day). Interval-contingent measurement is especially useful as it can collect relevant information at intervals of interest (e.g., sleep quantity or quality rating immediately after waking up) or can be used to gather anticipatory (e.g., ratings of how a spouse *thinks* he/she will feel during the upcoming day upon waking up) or global daily ratings (e.g., overall rating of satisfaction with marital interactions that day at bedtime).

Signal-contingent measurement protocols involve prompting participants to complete assessments at various times, typically in response to signals/cues delivered on a mobile device (e.g., palmtop computer, cell phone). The signaling can be scheduled at regular times or intervals, in an effort to collect data throughout the day (e.g., physical activity reporting every 15 min; Biddle et al. 2009). Signaled prompting can also occur at semi-random times throughout the assessment period. Semi-random prompts are less easily anticipated by respondents, thus reducing the likelihood that they will adjust their environment prior to the assessment. Examples of this approach are seen in several studies described previously that have assessed eating behavior across the day (Hilbert et al. 2008; Smyth et al. 2007).

Event-contingent measurement is used in cases when researchers are interested in specific events or experiences. For this strategy, participants are trained by investigators to self-initiate an assessment each time a specific event occurs. For example, in studies of children and adolescents, researchers have asked participants to complete EMA surveys following smoking a cigarette (Colvin and Mermelstein 2010) and after eating for children with a history of binge eating (Hilbert et al. 2008). This strategy is most reliant upon the effort and capabilities of participants, but is the best approach to ensure that events (particularly low-frequency events) are not "missed" (as they may be by an interval or signal contingent report).

As noted, multiple measurement schedules can be combined in a single study as well, and can leverage the unique strengths of each approach. For example, one may wish to carefully assess the experience of a target event, but supplement that (event) assessment with information collected early in the morning (interval) and at semi-random times throughout the day (signal). Colvin and Mermelstein (2010), for example, used a mixed approach in their study assessing adolescent smoking behavior, with participants both completing surveys at random intervals throughout the day (signal) as well as completing assessments after they smoked a cigarette (event). Combining measurement schedules in this way can thus provide researchers with information about participants' experiences both at times when they are and are not engaging in a behavior or event of interest. Determining the appropriate selection and combination of assessment schedules is, of course, guided by the research question(s).

Constructs to Assess: What Can Be Measured Using EMA?

Self-report data collection. As is likely evident by now, most EMA studies rely heavily (or exclusively) on self-report measures. In many cases self-report is the most efficient, and often the only, manner in which desired information can be collected. For example, there are few other options for obtaining information about individuals' internal experiences (e.g., mood, physical symptoms) or their perceptions of their social environment (e.g., other people's behavior). In other instances, however, it may be possible to collect data in peoples' everyday lives by using different real-world data capture modalities. These alternate methods for collecting real-time data can be broadly categorized as either *active* (i.e., require input from individual) or *passive* (i.e., do not require input from individual) data collection techniques. Each of these data collection strategies are briefly discussed below, with representative examples provided.

Active data collection (other than self-report). In addition to active self-report data collection, there are many other types of data that can be assessed through some active behavior on the part of a participant, such as measures of physical function (e.g., peak expiratory flow rate) or biomarkers (e.g., saliva samples). Although not the only example of non-self-report active collection methods, studies collecting ambulatory measures of salivary cortisol are becoming increasingly common and illustrate the information such techniques can provide. Cortisol is a hormone that, among its other properties, is important as it responds to stress, regulates some aspects of immune function, and is implicated in glucose metabolism. Ambulatory salivary cortisol is collected using small sterile tubes ("salivettes") that allow participants to provide saliva samples at various moments in daily life; these samples are then brought to a laboratory to assay biomarkers (e.g., cortisol). Slatcher and his colleagues have conducted several studies using ambulatory cortisol measures in studies of families. In a fascinating study examining the effect of stress on cortisol in marital relationships, husbands' and wives' provided ratings of stress, and saliva samples were collected to measure cortisol throughout the day. Results showed wives' cortisol levels were associated with their own work stress, but also with their husbands' work worries. These findings suggest that work stress affects not only one's own physiological stress levels (cortisol), but can also influence the physiological stress responses of one's partner (Slatcher et al. 2010). Although there is less research in children, in a second study Slatcher and Robles (2012) examined contextual and parental factors that may affect children's diurnal cortisol patterns by collecting six daily saliva samples (for 2 days) from children aged 3–5 living in two-parent families. Measures of conflict at home and parents' reports of child externalizing behaviors were also collected. Greater home conflict was associated with flatter diurnal cortisol slopes in the children, which has been tentatively related to negative health consequences in adulthood (e.g., Kumari et al. 2011). These studies illustrate ambulatory assessment measures that are alternatives to self-report, but still require active responses from study participants. More generally, these studies exemplify how active non-self-report measures may be used to inform our understanding of the dynamic relationship between real-world experiences and physiology in families.

Passive data collection techniques. Passive data collection techniques are becoming more commonly used as technology continues to advance, allowing for (often very large amounts of) data to be collected using ambulatory devices that do not require effort (or minimal effort) on the part of participants. For example, a variety of measures of physiological processes (e.g., blood pressure, heart rate, respiration rate, electrodermal activity) can be collected using small electronic devices that can be discretely worn on the body without significant interference with daily activities (Ebner-Priemer and Kubiak 2007; Patrick et al. 2005). Monitoring devices can take many forms and collect various physiological measures, including cuffs to measure blood pressure, watches, rings or pendants to monitor physical movements, and wireless chest straps or clothing designed to monitor heart rate, respiration, and/or blood glucose. These devices typically provide continuous monitoring without requiring any action by the user.

Although these tools are relatively new to family research, there is some evidence that they can be useful. For example, ambulatory blood pressure and heart rate have been tracked in a study of attachment orientation on physiological responses to real-world social interactions in adolescents (Gallo and Matthews 2006). In this study, high school students (aged 14–16) completed measures of anxious and avoidant attachment. They then wore an ambulatory blood pressure and heart rate monitor and tracked social experiences for 1.5 days. Results showed adolescents higher on anxious attachment exhibited higher blood pressure (both systolic and diastolic) when reporting current or recent interactions with friends. More avoidant adolescents had augmented ambulatory diastolic responses in response to real-world reports of social conflicts. This study also provides an example of how both active (i.e., self-report EMA) and passive (i.e., ambulatory blood pressure, heart rate) data collection techniques can be combined; as with signaling strategies, it is both common and often especially effective and informative to combine multiple methods within EMA.

In studies of adolescents, location and physical activity have also been tracked using passive assessment methods. For instance, in children aged 9–18 good convergent validity between ambulatory accelerometers used to measure physical activity (i.e., passive data) and self-reports of physical activity (i.e., active data collection) has been reported (Floro et al. 2009). The feasibility of using GPS-enabled cell phones to track adolescents movement has also been validated (Wiehe et al. 2008). In this study, the cell phones were used for both GPS tracking (passive) and to collect daily diary reports (active) from adolescents. Such studies are promising as they suggest GPS and other passive location and activity tracking devices could be useful for monitoring family members' movements in the real world.

It is also possible to collect information on the social environment. One such approach to doing so is the Electronically Activated Recorder (EAR), a digital voice recorder that records ambient sounds as people go about their daily lives (Mehl 2007). Participants wear the EAR device, which is programmed to periodically collect audio recordings, including conversations and ambient sounds; these recordings are then coded for a range of information about the social world (Mehl 2007). This assessment technique may be particularly useful in family research, as it provides a way to collect real-world information about contextual factors (e.g., social interactions, arguments, etc.) that may be difficult to collect via self-report. In a study described previously by Slatcher and Robles (2012), children wore the EAR, capturing ambient sounds while they were at home; these recordings were then coded for level of family conflict. This passive ambulatory assessment method is particularly useful in this example, as it was used with very young children (ages 3–5) who were not likely able to provide accurate self-reports of family conflict.

Understanding Moderators

Individual Difference Measures and EMA. Family researchers are often interested in understanding how person-level characteristics or individual differences influence individual and/or family experiences and dynamics. Momentary and daily assessment methods can be, and are, quite commonly incorporated into the study of person-level or family-level factors. For example, one might be interested in how an individual difference characteristic (e.g., psychological constructs, symptoms) or demographic variable (e.g., age, gender) influences EMA variables or moderates the relationship between variables assessed in everyday life.

In an EMA study of adolescents (aged 12–15) with varying levels of ADHD symptoms, Whalen et al. (2002) measured behavior, mood, and social contexts assessed in everyday life. Adolescents completed initial measures of their ADHD symptoms and then, via EMA, completed measures of behavior, mood, and social contexts every 30 min during two 4 day assessment periods. They found that teens with more baseline ADHD symptoms reported more negative mood, less positive mood, less alertness, more achievement-oriented pursuits, more time with friends and less time with family, and more tobacco and alcohol use. This example illustrates how characteristics of the individual can influence person-level processes in the real world. Conceptually, however, family or dyadic characteristics could also be used as either predictors or outcomes in similar analyses. For instance, EMA methods can be used to address questions regarding the manner in which family-level variables (such as family conflict, marital satisfaction, or cohesion) may influence real-world experiences or processes (e.g., momentary mood, coping, adjustment, etc.).

In a study of urban youth, Ewart et al. (2012) collected measures of motivation, everyday self-regulation, and cardiovascular risk. Motivational profiles were assessed in a structured interview format, everyday self-regulation was rated by the students' teachers, and cardiovascular functioning was assessed using ambulatory blood pressure monitors (programmed to record every 30–60 min for 2 days). Results showed that there was an effect of motivation style (in particular, agonistic striving, or the motivation to control other people's behaviors) on cardiovascular functioning, but this was moderated by emotion regulation. In particular, the relationship between agonistic striving and higher systolic blood pressure was stronger in youth rated as having lower everyday emotion regulation (Ewart et al. 2012). In addition to demonstrating how more stable individual difference variables may moderate momentary associations, this study provides an example of how EMA can also include multiple sources of data collection (e.g., youth interview, teacher ratings, passive physiological monitoring) in order to evaluate more complex relationships as they occur in everyday life.

Other Contextual Variables. In the family literature, there is also great interest in understanding how contextual variables, such as family and neighborhood environments, may influence individuals in their everyday lives. As was discussed previously, multi-system level designs are possible in which variables at various levels can be assessed, and data can be analyzed across levels to answer complex theoretical and applied questions. Just as person-level, individual difference factors (described in the previous section) can influence momentary within-person relationships, broader contextual variables (e.g., dyad, family, peer group, neighborhood, social context) can exert influence on within-person processes assessed via EMA. For example, Okon et al. (2003) have examined the effect of self-reported family-level contextual factors (i.e., family environment) on the daily experiences of young girls with bulimia nervosa. In this study, girls completed eight daily EMA surveys reporting on current family hassles and their bulimic symptoms. They also completed baseline measures of family conflict and emotional expressiveness. When girls reported experiencing family hassles, this was associated with bulimic symptoms later that day but only among girls who perceived their family as having high levels of conflict or low emotional expressiveness. No relationship between daily family hassles and subsequent bulimic symptoms was observed for girls who perceived to have normal family functioning. This study is an illustration of how using EMA, researchers can explore how family-level contextual factors (conflict, emotional expressiveness) may influence within-person relationships among constructs (momentary family hassles and symptoms).

This conceptual logic can be extended further—just as family-level factors may influence the relationship among momentary variables, the effect of contextual factors at "higher" levels (e.g., neighborhood, community) on family and/or individual processes can be examined. Furthermore, context can be evaluated from multiple other perspectives, not just from the individuals' perspective as described in the example above (Okon et al. 2003). For example, collecting information regarding contextual factors from objective sources and/or other reporters (e.g., family members, peers, teachers, trained observer) could help to provide measures of context more independent from a participant's self-reports.

Stable and Semi-stable Indicators. Although space precludes a detailed discussion of these issues, there are a range of relatively stable individual-level characteristics that researchers may wish to examine, particularly as they may relate to within person processes. For example, aspects of a person that could be viewed as stable might include genetic information and sex. Other aspects of a person may be "relatively" stable, such as a chronic disease diagnosis or attachment style. Some aspects are clearly dynamic, but change over intervals typically much greater/longer than those under exploration using EMA (i.e., they change, but over periods of months or years); these might include factors such as typical developmental processes (e.g., adolescence), marital status, obesity, and global well-being. Conceptually the logic for studying these variables is similar to that already discussed–these stable, or relatively stable, characteristics may serve to moderate processes both within and between individuals.

One example of this line of research is the study of how individual differences in genetics may be associated with health-related behaviors and processes (Way and Gurbaxani 2008). In an example of the potential of such approaches, Gunthert and colleagues (2007) assessed the serotonin transporter gene polymorphism (5-HTTLPR) and anxiety reactivity in daily life. College student participants provided salivary DNA samples (which were used to identify their allele length at 5-HTTLPR) and completed one daily web-based assessment of stressors and mood for two 30-day periods, separated by 1 year. Results showed that people with certain genetic makeup showed greater affective responses (more anxiety) on days when they reported experiencing more severe stress. Broadly speaking, this example illustrates the importance of considering how person-level variables (notably genetic variables, but also including other stable individual differences) may moderate dynamic relationships that are assessed in everyday life (in this case, also using a type of measurement burst design).

"New" Research Questions Addressed with EMA

Incorporating EMA methods into family research provides the opportunity to address complex and nuanced questions regarding family dynamics and processes in natural settings. In addition to the opportunities for research using intensive, real-world assessment methods described to this point, we would like to emphasize several priority areas where we believe EMA methods may be particularly fruitful for family research.

First, research questions that require cross-level moderation may be important to begin to explore. As has been discussed, studies that incorporate EMA methods to collect information about individuals can, and do, gather data from other "levels" of the family unit. Such designs allow researchers to ask questions about how factors at one "level" may influence variables or associations among variables at other "levels." Several examples of these types of questions and analyses have been discussed in detail in previous sections (e.g., parent or family-level factors affecting individuals). Such studies are critical to determine both the individuals at risk and contexts of risk, and to inform the design and implementation of interventions.

Second, as EMA methods collect multiple and often frequent assessments, these repeated measures can be used to compute a representative score, which can then be used in conjunction with other measures (such as standard self-report measures). Prior work has demonstrated that a self-report, aggregate measure–even those asking about reflections over a relatively short interval, such as a few days or 1 week–does not always correspond to data collected in real time. For example, Stone and colleagues have examined pain ratings in chronic pain patients using both EMA and retrospective methods. A comparison of the retrospective pain rating and the average EMA ratings during the same period revealed that participants reported significantly higher pain on the retrospective measure, as compared to the momentary EMA ratings (Stone et al. 2004). There is thus great research opportunity in determining what factors contribute to experienced reports (EMA) versus those that contribute to retrospective reports (standard measures). Moreover, how each may—perhaps differentially—relate to various outcomes.

Third, because EMA typically assesses people frequently, even over relatively short periods of time, this allows for the characterization of variability, particularly for processes that may fluctuate rapidly (e.g., over minutes, hours, or days; mood, behavior, etc.). Given this, EMA is uniquely useful for researchers interested in evaluating variability within persons, either as a predictor or as an outcome. This can allow researchers to begin to examine variability in real-world processes at either the group-level (e.g., family, dyad) or within the individual. For instance, emotional variability assessed via EMA has been shown to be associated with increased experience of pain and activity restriction among children with juvenile arthritis (Balabanis et al. 2012). EMA techniques–at least those with relatively high "density" of measurement–can thus provide the unique opportunity to measure processes that may fluctuate rapidly in everyday life and identify the predictors or consequences of such processes.

Fourth, although not unique to EMA, family researchers are often interested in collecting data from multiple perspectives in an effort to understand significant events and family experiences. EMA provides the opportunity to collect similar data, but to do so in naturalistic settings. This also allows for real time collection of multiple types of data and perhaps multiple informants (e.g., multiple family members) who are sharing the same environment and possibly experiencing the same "events" (e.g., a family meal, an argument, etc.). By having multiple EMA informants it is possible to represent each individual's perception of the "event" and, in particular, identify discrepancies in perception. This may have unique relevance for clinical intervention, particularly as it can be obtained in natural settings and also examined dynamically over time (through the use of the repeated EMA assessments).

Considerations for Implementing EMA Methods for Family Research

Although we are–not surprisingly–great advocates of EMA methodology, and see many benefits of applying EMA methods for the study of families and children, we acknowledge that this approach is not without challenges or limitations. EMA methods can be difficult and complex to implement, typically have high start-up costs, and require that a number of important decisions be made in advance of starting a study. For example, researchers must identify an appropriate measurement scheme (e.g., daily, within-day intensively; signal, interval, and/or event contingent), pick a device to use, obtain programming/implementation support, and consider measurement design (e.g., content, duration, etc.). Studies using EMA methods also often require more intensive participant training and monitoring in the field for problems (particularly those utilizing active reporting, such as self-report surveys). These issues are discussed in greater detail below.

Many EMA studies rely on the use of mobile electronic technology, such as mobile smart phones or tablet computers. These devices provide many benefits over paper-and-pencil diaries, including providing objective measures of compliance (by date and time-stamping assessments), eliminating the need for data entry, and providing a digital back-up of the data. EMA studies have successfully used mobile technology with children as young as 5 years old (Sherman et al. 2006) and with older adults as well (Sliwinski et al. 2009), but using these devices is not without its challenges. Researchers should take care to select both devices and survey software programs that are easy to use for the sample, and recognize it is often possible to modify the hardware, software, and/or study procedures to accommodate limitations of specific samples or individuals. For example, adaptations such as using a larger stylus or touch screen to respond on a smartphone for those who may have more difficult manipulating the devices, using larger size fonts, pictures to illustrate (or replace) survey questions, or even alternative technology with which to administer the study (e.g., voice recognition/recording to capture spoken responses) all may be appropriate adaptations for accommodating certain samples.

Taking part in EMA studies requires substantial time and effort on the part of the participants, as they must periodically take time out of their day for study activities. It is important to balance the frequency and duration of assessments used, in order to both appropriately address the research question and minimize participant burden. Studies using frequent assessments (e.g., self-reports every 30 min) should be relatively short in duration (e.g., seconds to minutes to complete) and occur over shorter assessment periods (i.e., several days at most). Alternatively, less frequent assessments (e.g., daily) can allow for respondents to complete longer assessments over greater periods of time (e.g., weeks to months). Although not unique to family research, careful consideration of the most appropriate periods during which to obtain assessments is also important for minimizing participant burden, while adequately collecting information during times or situations of interest. For example, in a study interested in parent-child interactions, only collecting assessments during times when parents and children are together (e.g., not while at school/work) may be appropriate. Another consideration when designing an EMA protocol is the development of items that are appropriate for the relatively short assessment timeframe required for EMA (i.e., momentary or very recent experiences). In studies using EMA with children, items that are theoretically and developmentally appropriate for the research question, study design, and children's age are necessary (see Sherman et al. 2006 for an example).

Given the typical reliance on active data collection in EMA (notably self-report), participant compliance is essential. Researchers can encourage compliance with the study protocol by ensuring participants are adequately trained (i.e., understand what they need to do), use hardware and software that allow for a user-friendly interface, and provide cuing or reminders (e.g., alarms). There is also evidence that, for both adults and children, using mobile electronic devices to collect EMA data (e.g., palmtop computers, smartphones) instead of paper diaries can significantly improve objective measures of compliance (e.g., Palermo et al. 2004). When using EMA with children, creating a sense of accountability (e.g., electronically tracking compliance), rewarding compliance, and using other family members to remind children to complete assessments can also help to improve compliance (see Hufford 2007 for a review of compliance with EMA protocols).

These challenges by no means should be taken to suggest that studies using daily assessment methods cannot, or should not, be used with children or families. In fact, with the appropriate adaptations and training of participants, EMA data collection methods are generally rated as acceptable (i.e., easy to use, minimally disruptive to daily schedule) by respondents (e.g., Palermo et al. 2004). There is, however, no simple "off the shelf" implementation strategy for EMA assessment methodology. Great care must be given to design and implement an EMA strategy unique to the research question and sample at hand.

Ecological Momentary Interventions: Integrating EMA with Interventions

Although EMA was developed for (and is typically used as) an assessment tool, there is emerging interesting in using real-time assessment and intervention methods as part of treatment programs. Integrating real-time data capture into intervention tailoring and/or delivery is referred to as Ecological Momentary Interventions (EMI). EMI are characterized by the delivery of interventions (or intervention components) to people as they go about their daily lives, and thus occur in the natural environment and are delivered at specific moments in everyday life (Heron and Smyth 2010; Patrick et al. 2005). To date, mobile technology has been used to deliver interventions in patients' everyday lives for a variety of conditions and health behaviors including, but not limited to, anxiety, weight loss, smoking cessation, and alcohol use (for a review, see Heron and Smyth 2010).

An emerging area is the development of individually content and time tailored interventions that are delivered in real time. Although tailoring the content of interventions to stable participant characteristics is fairly commonplace, EMI protocols allow for the timing of the delivery of EMI to be individually tailored as well. For example, in a smoking cessation program, EMI can be delivered as text messages to participants cell phones at specific times when they typically smoke (e.g., Rodgers et al. 2005). EMI can also be implemented using people's recent or concurrent EMA reports. Studies wherein participants periodically complete EMA regarding their current affective state (e.g., anxiety level) or behaviors (e.g., calories consumed) have been provided EMI only at times when they were reporting risky behaviors or experiences (Burnett et al. 1992; Newman et al. 1999). Similarly, using "passive" ambulatory assessment devices, people's momentary physiological states and environmental conditions could also be used to trigger and tailor real-time interventions. EMI based on time-tailored treatment approaches appear to be especially useful for behaviors or conditions with distinct and recognizable antecedent states (e.g., cravings/urges, negative emotions) or events (e.g., target stressors, mealtimes, social situations) for which EMA assessments and corresponding intervention components can be developed. To date, however, research testing whether such delivery systems are more efficacious than traditional formats is sparse, and this remains a priority area for future research.

There may also be unique opportunities in family research. Although we are not aware of any existing research in this area, family-based EMI appear to hold tremendous theoretical and clinical promise. If multiple members of a family are each completing an EMA protocol, this opens up interesting assessment and intervention possibilities. For example, assessment of both participants in a dyadic experience (e.g., a spousal interaction) might reveal unique antecedents to subsequent problems (e.g., a discrepancy between the two participants' ratings of the interaction—such as how pleasant it was, if it was an argument or not, etc.—may uniquely predict later outcomes in ways that the report of either member would not). Another approach might be inter-individual EMI; that is, using the risk cues from one person to provide

EMI to not only that person, but perhaps others as well. Examples might include occurrences of child poor behavior leading to parent-targeted EMI to promote better interactions, or when one member of a dyad is experiencing high stress not only that individual receives a (stress-management) EMI, but their partner may receive some EMI as well (informational, behavioral, etc.). We recognize the theoretical, practical, and ethical difficulties of designing and implementing such interventions; these examples are meant merely to provoke thought, and are not presented as fully formed or specific suggestions.

Conclusions

This chapter illustrates the potential benefits of EMA approaches to family research. Notably these benefits include, but are not limited to, the reduction of reporting biases, enhanced ecological validity, and the exploration of dynamic and time-related processes and variability within individuals. Although self-report data may continue to serve as the "core" of EMA approaches, there is an increasing range of additional active and passive assessment systems that capture other data streams in real time (e.g., location, social behaviors, biomarkers, physical activity, etc.). EMA methodologies can also facilitate researchers building models across relatively time invariant (e.g., sex, genetic) and time varying (e.g., affective, social) processes, and across multiple levels of assessment and analysis (e.g., within and between person, short and long-term temporal processes, and nested in different family, social, and cultural contexts). Finally, there is also great potential for extending linkages between EMA methods and intervention research. Given the nature of the theoretical and practical issues addressed by family researchers, we believe that EMA may prove a useful additional tool; one that provides a unique opportunity for the detailed study of how family and individual biopsychosocial processes unfold in time and in context.

References

Balabanis, M., Bromberg, M., Anthony, K. K., Gil, K. M., Franks, L., & Schanberg, L. E. (2012). Emotion regulation predicts pain and functioning in children with juvenile idiopathic arthritis: An electronic diary study. *Journal of Pediatric Psychology, 37,* 43–52.

Biddle, S. J. H., Gorely, T., & Marshall, S. J. (2009). Is television viewing a suitable marker of sedentary behavior in young people? *Annals of Behavioral Medicine, 38,* 147–153.

Burnett, K., Taylor, C., & Agras, W. (1992). Ambulatory computer-assisted behavior therapy for obesity: An empirical model for examining behavioral correlates of treatment outcome. *Computers in Human Behavior, 8,* 239–248.

Colvin, P. J., & Mermelstein, R. J. (2010). Adolescents' smoking outcome expectancies and acute emotional responses following smoking. *Nicotine & Tobacco Research, 12,* 1203–1210.

Connelly, M., & Bickel, J. (2011). An electronic daily diary process study of stress and health behavior triggers of primary headaches in children. *Journal of Pediatric Psychology, 36,* 852–862.

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience sampling method. *Journal of Nervous and Mental Disease, 175,* 526–536.

Ebner-Priemer, U., & Kubiak, T. (2007). Psychological and psychophysiological ambulatory monitoring: A review of hardware and software solutions. *European Journal of Psychological Assessment, 23,* 214–226.

Ewart, C. K., Elder, G. J., & Smyth, J. M. (2012). How implicit motives and everyday self-regulatory abilities shape cardiovascular risk in youth. *Annals of Behavioral Medicine, 43,* 286–298.

Floro, J. N., Dunton, G. F., & Delfino, R. J. (2009). Assessing physical activity in children with asthma: Convergent validity between accelerometer and electronic diary data. *Research Quarterly for Exercise and Sport, 80,* 153–163.

Gallo, L. C., & Matthews, K. A. (2006). Adolescents' attachment orientation influences ambulatory blood pressure responses to everyday social interactions. *Psychosomatic Medicine, 68,* 253–261.

Gunthert, K. C., Conner, T. S., Armeli, S., Tennen, H., Covault, J., & Kranzler, H. R. (2007). Serotonin transporter gene polymorphism (5-HTTLPR) and anxiety reactivity in daily life: A daily process approach to gene-environment interaction. *Psychosomatic Medicine, 69,* 762–768.

Heron, K. E., & Smyth, J. M. (2010). Ecological Momentary Interventions: Incorporating mobile technology into psychosocial and health behaviour treatments. *British Journal of Health Psychology, 15,* 1–39.

Hilbert, A., Rief, W., Tuschen-Caffier, B., de Zwaan, M., & Czaja, J. (2008). Loss of control eating and psychological maintenance in children: An ecological momentary assessment study. *Behavior Research and Therapy, 47,* 26–33.

Hufford, M. R. (2007). Special methodological challenges and opportunities in ecological momentary assessment. In A. A. Stone, S. Shiffman, A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research* (pp. 54–75). New York: Oxford University Press.

Kumari, M., Shipley, M., Stafford, M., & Kivimaki, M. (2011). Association of diurnal patterns in salivary cortisol with all-cause and cardiovascular mortality: Findings from the Whitehall II study. *The Journal of Clinical Endocrinology and Metabolism, 96,* 1478–1485.

Mehl, M. R. (2007). Eavesdropping on health: A naturalistic observation approach for social-health research. *Social and Personality Psychology Compass, 1,* 359–380.

Newman, M., Consoli, A., & Taylor, C. (1999). A palmtop computer program for the treatment of generalized anxiety disorder. *Behavior Modification, 23,* 597–619.

Okon, D. M., Greene, A. L., & Smith, J. E. (2003). Family interactions predict intraindividual symptom variation for adolescents with bulimia. *International Journal of Eating Disorders, 34,* 450–457.

Palermo, T. M., Valenzuela, D., & Stork, P. P. (2004). A randomized trial of electronic versus paper pain diaries in children: Impact on compliance, accuracy, and acceptability. *Pain, 107,* 213–219.

Patrick, K., Intille, S., & Zabinski, M. (2005). An ecological framework for cancer communication: Implications for research. *Journal of Medical Internet Research, 7,* e23.

Pickering, T., & Freedman, R. (1991). The white coat effect: A neglected role for behavioral factors in hypertension. In P. McCabe & N. Schneiderman (Eds.), *Stress, coping, and diseases* (pp. 35–49). Hillsdale: Laurence Erlbaum Associates Inc.

Rodgers, A., Corbett, T., Bramley, D., Riddell, T., Wills, M., Lin, R.-B., & Jones, M. (2005). Do u smoke after txt? Results of a randomised trial of smoking cessation using mobile phone text messaging. *Tobacco Control, 14,* 255–261.

Scollon, C. N., Kim-Prieto, C., & Diener, E. (2003). Experience Sampling: Promises and pitfalls, strengths and weaknesses. *Journal of Happiness Studies, 4,* 5–34.

Sherman, S. A., Eisen, S., Burwinkle, T. M., & Varni, J. W. (2006). The PedsQL Present Functioning Visual Analogue Scales: Preliminary reliability and validity. *Health and Quality of Life Outcomes, 4,* 75.

Slatcher, R. B., & Robles, T. F. (2012). Preschoolers' everyday conflict at home and diurnal cortisol patterns. *Health Psychology, 31*(6), 834–838.

Slatcher, R. B., Robles, T. F., Repetti, R. L., & Fellows, M. D. (2010). Momentary work worries, marital disclosure, and salivary cortisol among parents of young children. *Psychosomatic Medicine, 72,* 887–896.

Sliwinski, M. J. (2008). Measurement-burst designs for social health research. *Social and Personality Psychology Compass, 2,* 245–261.

Sliwinski, M. J., Almeida, D. M., Smyth, J., & Stawski, R. S. (2009). Intraindividual change and variability in daily stress processes: Findings from two measurement-burst diary studies. *Psychology and Aging, 24,* 828–840.

Smyth, J. M., & Stone, A. A. (2003). Ecological Momentary Assessment research in behavioral medicine. *Journal of Happiness Studies, 4,* 35–52.

Smyth, J. M., Wonderlich, S. A., Heron, K. E., Sliwinski, M. J., Crosby, R. D., Mitchell, J. E., & Engel, S. G. (2007). Daily and momentary mood and stress are associated with binge eating and vomiting in bulimia nervosa patients in the natural environment. *Journal of Consulting and Clinical Psychology, 75,* 629–638.

Stone, A. A., Broderick, J. E., Shiffman, S. S., & Schwartz, J. E. (2004). Understanding recall of weekly pain from a momentary assessment perspective: Absolute agreement, between- and within-person consistency, and judged change in weekly pain. *Pain, 107,* 61–69.

Stone, A. A., & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine, 16,* 199–202.

Stone, A. A., Shiffman, S., Atienza, A., & Nebeling, L. (2007). Historical roots and rationale of ecological momentary assessment (EMA). In A. A. Stone, S. Shiffman, A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research* (pp. 3–10). New York: Oxford University Press.

Way, B. M., & Gurbaxani, B. M. (2008). A genetics primer for social health research. *Social and Personality Psychology Compass, 2,* 785–816.

Whalen, C. K., Jamner, L. D., Henker, B., Delfino, R. J., & Lozano, J. M. (2002). The ADHD spectrum and everyday life: Experience sampling of adolescent moods, activities, smoking, and drinking. *Child Development, 73,* 209–227.

Whalen, C. K., Henker, B., Jamner, L. D., Ishikawa, S. S., Floro, J. N., Delfino, R. J., … Johnston, J. A. (2006). Toward mapping daily challenges of living with ADHD: Maternal and child perspectives using electronic diaries. *Journal of Abnormal Child Psychology, 34,* 115–130.

Wiehe, S. E., Carroll, A. E., Liu, G. C., Haberkorn, K. L., Hoch, S. C., Wilson, J. S., & Fortenberry, J. D. (2008). Using GPS-enabled cell phones to track the travel patterns of adolescents. *International Journal of Health Geographics, 7,* 22.

Chapter 10
Why Qualitative and Ethnographic Methods Are Essential for Understanding Family Life

Thomas S. Weisner

Qualitative and ethnographic research methods are essential for understanding family life. Qualitative and case study materials have been staples for family research throughout the social sciences from the beginning. Family histories, cultural contexts, everyday routines and practices, narratives, experiences, intentions, stories, triumphs, secrets, troubles, and pain all matter deeply, and are what families mean to us. This information surely deserves to be understood and used in our research. Without incorporating qualitative methods in family research, those aspects of family life can never be fully captured.

For very good *analytic* research reasons we want to include quantitative and other methods and systematic research designs. We make the necessary analytic assumptions, pretending that the world is linear and additive and predictable, that we can bracket out context, and that our analytic categories actually match the way families and households are constituted (Weisner and Duncan, in press). But the worlds of families are not linear and additive, and context matters profoundly. Qualitative methods, ethnography, and fieldwork provide essential ways to include rich detailed information found in family narrative. Thus, any concern over the use of qualitative methods certainly should not be whether such evidence can be valuable; it already is. The question rather is how best to collect such information in ways that are productive, meaningful, believable, and add value to research. In this chapter, I provide some examples of such research, suggest ways to assess the quality of qualitative work, and emphasize the value of integrating qualitative evidence with quantitative data.

Most theoretical frameworks in family research are open to qualitative evidence. For family research as for so many topics, *combining* biological substrate, ecological setting, beliefs and behaviors, and the experiences and meaning systems of individuals in families, all then followed through developmental time, is our widely accepted conceptual and heuristic framework. Family systems approaches, including Bronfenbrenner's model that blends "person, process, context and time" (Bronfenbrenner

T. S. Weisner (✉)
Departments of Psychiatry and Anthropology, University of California, Los Angeles, CA, USA
e-mail: tweisner@ucla.edu

S. M. McHale et al. (eds.), *Emerging Methods in Family Research,*
National Symposium on Family Issues 4, DOI 10.1007/978-3-319-01562-0_10,

1995, 2005), provide conceptual frameworks that invite qualitative methods. Qualitative and ethnographic methods provide information on settings and contexts, and on the experiences, meaning systems and normative scripts that drive family life and *direct* our behavior (D'Andrade and Strauss 1992).

Qualitative Methods: Epistemology and Integration with Other Methods

Qualitative understanding is not inherently incommensurate with other methods. Qualitative methods are very different, for sure, and might well be preferred over quantitative methods for understanding meanings, experience, interpretations, intentions, cultural models and scripts, and narratives and stories family members have about their world. Any method can be the preferred or best for addressing certain topics—for representing those topics effectively—but other methods can add value. A representational or correspondence approach to methods seems most useful; all methods attempt to describe and to represent, and claim they correspond to the world. Thus:

> . . . quantitative research [provides methods] of inquiry that analyze numeric representations of the world. Survey and questionnaire data as well as biological or physiological data are often analyzed in quantitative units. Inquiry that relies on qualitative methods collects and analyzes non-numeric representations of the world—words, texts, narratives, pictures, and/or observations. The epistemological assumption . . . is that in scientific endeavors, the world can be represented through both numbers and words and that numbers and words should be given equal status in [family research]. (Yoshikawa et al. 2008, p. 344)

There are ways to characterize qualitative methods other than to contrast them with quantitative methods, even though this dichotomy, and all that is associated with it, is the natural language paired opposite terminology. Hence, researchers are so often trained either in the "qualitative track" or "quantitative track" for methods, rather than the mixed methods track or the "narrative plus statistical track". For instance,

> Anthropologists have described methods as experience-near (representing the voices, intentions, meanings, and local rationality of parents and children in local settings) and experience-distant (representing the world of groups, institutions, and social address categories). Methods can be particularistic, capturing only a part of some phenomenon, or holistic, attempting to capture the whole context or situation . . . (Yoshikawa et al. 2008, p. 345).

Family researchers usually become specialists in specific methods and aspects of family life, but hopefully they do not become methodocentric. *Methodocentrism* is identification with or commitment to certain methods turned into a personal identity or ideology as opposed to considering methods as tools for representing the family topics we are trying to understand (Weisner 1996). ("I am a qualitative person; I defend such methods as unique and as more valuable and valid, and reject others," can be a methodocentric position.) Methodocentrism can lead to confusing the topics about family life we want to study (e.g., attachment, sibling relationships, family budgeting, work-family balance) with particular methods for studying them (e.g., the Strange Situation, questionnaires, daily routine diaries, stress scales).

Donald Campbell (Webb et al. 1981) argued long ago that the reason for using multiple methods is that all methods are weak and are only partial indicators of the underlying events in the world we want to describe. We need qualitative methods alongside others because qualitative methods have strengths and weaknesses, just as do questionnaires and surveys. The weaknesses of naturalistic participant observations of family life or open-ended conversational interviews are complemented by the strengths of closed-ended surveys and 5-point scales on a questionnaire, *and vice versa*. A pluralist position with regard to methods does not in principle privilege one way of representing the world (numbers, or models) over another (narratives, text, fieldwork experiences). Of course some kinds of phenomena in the world are best represented by narrative experiences while others are best represented by numbers or models.

I do not want to give the impression that this expansive, positive view of qualitative methods is by any means universally agreed on, that there are no "paradigm wars", or that there is not a great deal of remaining ambiguity and interpretive work required for many kinds of ethnographic and qualitative methods in family research and the social sciences. In anthropology for example, qualitative methods, ethnography, and fieldwork seemingly are continually in a crisis of representation, practice, and meaning. Borneman and Hammoudi (2009) provide a recent overview of the variety of approaches to qualitative field research in anthropology, including "... anthropoesis, dialogism, genealogies of modernity, history, world system, transnationalism, autoethnography, the staging of multiple voices, science studies, simple activism, and critiques of knowledge through the study of constructed subjectivities" (p. 4).

Any research encounter involving fieldwork and a personal relationship with participants will be fraught with reflexivity, contingencies, unknown method effects, and situational influences. Every element of that method potentially then can be contextualized and critiqued.

> ... fieldwork is the registering of sensory impressions in a (temporal) process of mutual subject-discovery and critique, an engagement with persons, groups, and scenes that takes into account the dynamics of our interactions as well as the differences between our locations and those of our interlocutors. (Borneman and Hammoudi 2009, p. 19)

The analysis of the fieldwork encounter itself then becomes a part of our interpretation of the evidence gathered through such encounters.

The reasons why qualitative fieldwork can be so highly contested ironically often are the same reasons that qualitative fieldwork is seen as essential for social science. The family and social world is of course constantly changing, unstable, and globalizing. Family life has multiple and contested meanings and embodies history (an intimate family history and the broader social, national, immigrant, and economic histories). The family and community units chosen to describe families themselves (e.g., nuclear, legally married) can predetermine our descriptions, analyses, and results and clearly do not fully represent this changing complex family world. Therefore, fixed quantitative methods and categories can not possibly fully capture such family worlds today, and so we need more open, fluid, context-examined qualitative methods.

So qualitative inquiry and fieldwork involves some risk and uncertainty and requires some degree of openness. The method depends on the high likelihood that by ceding to our family participants some control over the research setting and collection of data, new evidence and understanding of family and community life will result in large part because of the openness, greater disclosure and shared ownership of the research partnership. Relations of greater dialogue and trust can and do then emerge in qualitative research and valuable information results.

Assessing Qualitative Research

Criteria for assessing qualitative work exist and are used, though there certainly is less agreement on the criteria and on how to use them in contrast to assessing quantitative methods. As when assessing quantitative methods, a clear conceptual framework, sample description, design, participant consent, and evidence of effective study implementation can be used to assess qualitative and fieldwork studies. In addition, qualitative methods can effectively be assessed for their depth, richness and complexity of descriptions, breadth, ability to move across levels of analysis, veridicality (the use of specific exemplars and vignettes illustrating the topics and findings of interest), and holism. Incorporating the terms and concepts used by participants themselves, including their own explanatory models, reasons and motives offered to account for their actions are also important in qualitative data (Weisner 1996).

There are useful framing or checklist criteria for assessing studies that integrate qualitative and quantitative data (Greene 2007; Lieber 2009; Small 2011; Weisner 2005). Weisner and Fiese (2011) suggest these questions to ask of such research:

> Is the rationale for a mixed methods approach clearly specified? Are the qualitative and quantitative data systematically integrated in such a way that maximizes the strengths and minimizes the weaknesses of each approach? Is the form of data integration clearly specified? Do the authors clearly identify how they integrate quantitative and qualitative data either through merging, connecting, or embedding data (Creswell et al. 2011)? At what phase of the study was the mixed methods approach introduced (e.g., pilot phase, program evaluation, embedded in longitudinal study)? Is the method of data collection clearly specified for both the quantitative and qualitative data? If interviews or video recorded data were used, how were the questions or video samples derived? How were the qualitative data reduced and summarized? If the qualitative data were coded, how were the coders trained? What were the rules for transcription? How was consensus reached? If the report is part of a larger study (as many mixed method studies are embedded in larger studies), how is it distinct from other published reports or those under review? Do the textual or narrative examples provide sufficient detail (without being redundant) to illustrate key findings? Does the discussion highlight the advantages and limitations of a mixed methods approach? (p. 797)

Of course no study can nor has to include every element described above; studies are specific and can never account for all methodological challenges or contingencies. However, the point is that there are useful frameworks available for asking about the reliability, validity, and believability of qualitative work, and these are being used.

The Suite of Qualitative, Ethnographic and Fieldwork Methods

Qualitative methods in family research often are thought of as consisting primarily of open-ended questions with probes in conversational interviews. Such interview methods are important but are certainly not the only qualitative method. Indeed, a large suite of approaches has been developed to represent family life. These include community participation and ethnographic observations (Bernard 2013). Ethnography in family research is the account of the way of life of a family or community using interviews and field participant observations. Fieldwork involves observing, talking with, and perhaps participating in the everyday lives of family members, often across settings beyond the household. In addition to narrative texts, qualitative work includes pictures, video, found objects, and observational methods. There are many kinds of participant and systematic observational methods, and many types of interviewing (informal conversations, guided conversations, use of probes, focus groups, and others). Gilgun (2012) identifies four characteristics of qualitative family research traditions in social welfare including the use of qualitative methods to capture experiences in context, extensive direct engagement with the family, interpretations grounded in individuals' accounts of their experiences, and research that looks toward promoting change and action.

How qualitative data are subsequently *analyzed* can be very different from how they were collected. Fieldnotes, interviews, and naturalistic video frequently are coded for quantitative analysis using standard criteria for reliability and validity. *Indexing* or "bucketing" is not the same as weighted coding with reliability tests. Indexing is used to mark off a long interview or excerpts from notes according to general topics (content related to health, or siblings, or academics or couple relationships and conflict) for analysis. The qualitative analytic process of "*structured discovery*" is common in qualitative and mixed methods work, during which the methods and subsequent analytic strategies remain open to unexpected processes and patterns yet focus on project-specific topics such as parenting, experiences with welfare systems, or family routines. "Grounded theory" is another analytic approach commonly described as a way to explore patterns by close, iterative listening, reading, and observing of the sample data (LaRossa 2005; Strauss and Corbin 1998).

Numerical as well as text data on family processes frequently co-exist within the same original qualitative study. Harkness et al. (2011) for example, reported on five family samples in five European and US communities focused on family time, meal times in families, play, and school-related or academic-promoting activities. The cultural meaning of these activities was important for understanding the time spent together by families. Interview text (the qualitative narrative accounts parents used to explain why they organized their routines as they did) was quantitatively scored, and typical days (quantitative data drawn from diaries and counts) were qualitatively summarized based on patterns found in the diary and questionnaire data. This analysis illustrated, as the authors put it, that "... qualities can be counted, and quantities can be described" (Harkness et al. 2011, p. 811).

Uses of Qualitative and Ethnographic Fieldwork Methods in Family Studies

Gaining participant trust and rapport. One use of qualitative inquiry is very common: gaining greater levels of trust, familiarity, and rapport between researcher and participant. This closer relationship enhances data quality, engagement, and retention of research participants. Another technique is explicitly and actively bringing the research participant/subject into the *shared* data collection project. For example, our current La Vida family study has followed over four hundred 14- to 16-year-old Mexican–American adolescents in Los Angeles for two years (Andrew Fuligni, PI, Nancy Gonzalez and Thomas Weisner, co-PIs). The teens and parents complete daily diaries, questionnaires, and a structured survey interview. We also gather school records. In addition, 10 % of La Vida families, randomly sampled from the full study sample, agreed to be part of a qualitative and ethnographic study. They were interviewed in their homes using the Ecocultural Family Interview, a conversational interview with prompts and probes.

One may wonder how we engaged 14- to 16-year-old adolescents in an extensive, sometimes personal conversation about their lives for the first time and receive rich and informative accounts of their lives. (We actually obtain responses beyond "yeah; whatever; I dunno ...") In a visit prior to the interview visit our team gave the teens digital cameras and asked them to take 25 photos of people, places, objects, activities, or whatever was important to them in their lives. When we arrived for the interviews a few weeks later, we plugged those cameras into our laptops and started talking with the adolescents about the photos. Who is that? Oh, your girlfriend? Teacher? For which class? That's your soccer team.... That's your room, favorite music group poster, your Mom cooking dinner. What chores does she do and what do you do? Teens took photos *of other family photos*: their relatives in Mexico they could not see and missed. They took pictures of their small home shrines to saints or the Madonna, their churches, places they wanted to work someday (police station; restaurant; hospital; offices). The range of important family information embedded in those photos was often surprising and remarkable, and the engagement of the teens in talking with us about a wide range of topics was far greater than would have been true otherwise.

We did not have a sample of teens who were not asked to take photos, and we did not ask teens specifically about the value of photos in sharing information. No doubt, many features contributed to the usefulness of the photos: The teens were given an active, agentic role in the research process; nonverbal visual communication was comfortable for many; personal experiences could be indirectly captured in photos; people, contexts, and relationships could become a part of the narrative account and told a meaningful story; and the research questions about family life in our project often led off of the photo narratives. We repeated this process again a year later during our second qualitative home visit. We currently have an archive of over 1,000 photos linked to interviews that have been coded and linked to quantitative data on the adolescents and their family daily routines.

Our interview topics usually emerged simply by extending conversations from the photos. This study also used “show cards”—large-print laminated cards with phrases (e.g., family rules and responsibilities, daily routine, time together, trust and hidden activities, school, religion, future goals, work/income/money, financial stress, friends-peers-family, your family story/history) which we set out on the table or couch, reminding the interviewer and teen or parent of the topics to be discussed. Qualitative interviews are very open and conversational, but this does not mean they cannot be made more comparable. For example, we minimize false negatives in qualitative interviews by always bringing up core topics if the parent or teen does not bring them up themselves. The La Vida taped interviews are transcribed (and translated if in Spanish) and uploaded to web-based qualitative analysis software named Dedoose (Dedoose.com) for indexing (indexing and coding done through drag and drop code trees placed onto highlighted text), as well as reliable quantitative coding (reliability tests are automated within a Training Center in Dedoose, for example). These data are then linked to quantitative data (i.e., school achievement, survey and questionnaire scale summary scores, demographic data) in addition to the qualitative summary of patterns in the narratives (Steinberg 2012). Coded and indexed interview text, numerical summaries of code-by-code matrices, and Excel spreadsheets with quantitative summaries of codes and other quantitative measures, including charts and graphs, can be exported directly from software such as Dedoose, into Word or Excel for further analysis and inclusion in papers.

Unpacking analytic categories; discovering new ones. Another important use of qualitative and fieldwork methods is to unpack standard social address categories (conjugal, single mother, dual earner married, extended). Qualitative and ethnographic researchers use existing family categories, but researchers are in a better position to *question* the categories as they are closer to hearing and seeing who is in the home and why, what roles are played, and what happens over time (Roy et al. 2008). Qualitative methods allow for the study of the motivations, strategies, and intentions of family members themselves with regard to finding support, forming alliances, establishing co-residence, marrying, and so forth. Such work, at ground level so to speak, can then actually lead to the creation of new categories for more systematic study.

The household and relationship category of “living together apart (LTA)” is an example of a new discovery emerging from qualitative research. Frequent separations followed by reunifications and cohabitation of a couple along with (either partner’s) children can create an LTA family unit. LTA households and families are very widespread in the US and throughout the world—yet there is no category for them conventionally available in census or survey work (Cross-Barnet et al. 2011). Of course family disruptions due to death, divorce, migration or many other reasons lead to episodic family and household formation, dissolution and reconstitutions. Cross-Barnet and colleagues identified a wide range of household formations, including “. . . stable marriages or cohabitations, serial cohabitations, intermittent cohabitations, LTAs, and abated unions (in which the mother does not engage in any romantic, sexual, or cohabiting relationship). . . .” (pp. 637–638). Each of these family formations was uncovered through close qualitative documentation of family life.

Interpretation of experimental and intervention study findings using qualitative evidence. Both qualitative and ethnographic evidence were important for the New Hope study and illustrate how such methods can be used in family and intervention research, and to better understand studies with experimental designs. New Hope (NH) was a successful poverty reduction program that offered a positive social contract to working-poor adults to support them and their families (Duncan et al. 2007; Yoshikawa et al. 2006). New Hope operated as a community organization in Milwaukee, Wisconsin for several years, from the mid- to-late 1990s. Participants who worked full time were eligible to receive significant income supplements (increasing income from the current level to 200 % of the federal poverty level), childcare vouchers which could be used at any licensed family care or center care facility, and health care benefits. Participants who did not have a job were given a community service job. All were shown client respect when they came to the NH office. Over 1300 adults in two low-income neighborhoods signed up for a lottery to participate in NH; half were randomly assigned to participate in the program, and half were not. Program and control participants and their families were followed for eight years from the time of their entry into NH. Surveys, questionnaires, teacher reports, and administrative data were all used to follow participants in the family study with children ages 1–12.

In addition, some 8 % of program and control families were randomly selected to participate in a qualitative and ethnographic study. They were visited multiple times in their homes for conversational interviews focused on the study topics and their lives and experiences. Topics included use of program benefits or use of other programs if in the control group, parenting, child care choices, work, marriage and partners, budgeting and income, substance use, religion, experiences of discrimination, education issues, and others (e.g., *tell us about birthdays, holidays*; *if you could talk to NH and other parents about your experiences, what would you say*). We listened to their stories in their own words, grounded in their own contexts and life experience. Our qualitative teams followed families in this subset for roughly six years. We summarized these interviews and home observations by topics and systematically coded or indexed them using web-based online software for team use. Both the coded qualitative data and the patterns of narratives from the interviews and notes were analyzed (Duncan et al. 2007; Weisner and Duncan, in press; Gibson and Weisner 2002; Weisner 2011b; Yoshikawa et al. 2006).

Quantitative and qualitative evidence in combination provided valuable information as the NH intervention played out in family lives. The qualitative sample suggested motives, strategies, and family circumstances which helped us understand what turned out to be selective take up of benefits, for example (Gibson and Weisner 2002). Developmental data showed that boys in program families benefited academically and in classroom behavior reports by teachers, relative to boys in the control families, while girls did not. This was a surprising result not easily explained from quantitative data. Some parents in the NH sample had a greater likelihood of marrying or finding stable partners than the control group. Women described how they sometimes got out of bad existing relationships, finally got their own housing and found some stability in their lives, thus making new relationships possible.

Qualitative evidence helped interpret results from quantitative data and treatment-control impact analyses (why boys did better, why some work trajectories were more successful, how subjects found new and often better partners, what was behind the selective take up of NH, how parents talked about their family—work balance choices and intentions, and how parents balanced supports they needed from kin and others, with the obligations and risks such supports also entailed). In other cases, qualitative evidence simply stood on its own as rich knowledge about the experiences and lives of working poor parents and their kids (their own definitions of "being a good mom", immigrant and migration stories, discrimination stories, the very strong importance of religion for some, and domestic violence situations). This is, at the end of the day, the reason for using qualitative methods in family research: Qualitative methods are core to and add very significant value to scientific understanding. This is the same standard as any method should be held to, and qualitative and ethnographic methods can more than meet this standard.

Describing processes of family change over time. Family life always includes fluid processes emerging over time, where there is strategic and tactical intention by family members. Qualitative methods are very effective at capturing these processes. "Kinscription" for example, is a family process term that came from qualitative evidence. Kinscription describes the constant attempts by single mothers to recruit and involve biological fathers, and other romantic/intimate partners and their kin, to help their families and be involved in their and their children's lives. Kinscription processes are central to the lives of millions of mothers in the United States and elsewhere (Roy and Burton 2007). They accounted for some of the family processes and household formation among New Hope working poor families as well.

Identifying holistic patterns and themes. Qualitative methods also can suggest profiles or ideal types for family processes that cut across demographic or measurement/scale categories. For example, Lareau (2003) described two broad prototypes or clusters of class differences in how parents prepared their children to adapt to their schools, neighborhoods, and family prospects: middle class "concerted cultivation" contrasted with lower class or working poor "natural growth" models for socialization and child development (cf. Kohn 1977). These profiles incorporate earlier qualitative and mixed methods studies of class differences in discipline, language use, stimulation, parenting styles and investment.

Interpreting, contextualizing, challenging quantitative data. Qualitative evidence helps inform us what the number "3" circled on a questionnaire means; what were our informants thinking about that item or scale topic; and why did a subject answer "yes" to a survey question. An interesting teaching exercise (or research method) asks a group to answer common items on well-known scales from 1–n, and then ask *why* someone chose 3, another 5, another 2 and so forth. Here are some items from a familism scale, for example: How important is it to you that ... [your child] treat his/her parents with great respect?; [your child] live or go to college near his/her parents?; [your child] make sacrifices for your family? (1 = *not at all importan,* 2/3 = *somewhat*, 4/5 = *very important*). The explanations and even very brief follow-up questions and answers are often highly variable, interesting in their own ways, and revealing far beyond the circled number. For example: What things indicate respect; why is it very important? What sacrifices have you have made? Why is it important to go far away for college? Revealing differences and class discussions emerge.

Qualitative data can be essential for interpreting and contextualizing survey and questionnaire data. Pearce and Denton (2011) used qualitative narratives about religion from teens, along with a quantitative analysis of five patterns of religious life (named Abiders, Adapters, Assenters, Avoiders, Atheists). Pearce and Denton preferred a person-centered approach to the *interpretation* of their data, over a variable-centered approach. "... youth see distinctly the multiple dimensions of religion and are comfortable packaging them together in various ways, even when their intensity or importance is not always consistent" (Pearce and Denton 2011, p. 140). The evidence from their study suggested that there isn't a unilineal religiosity scale that represents US adolescents along a simple high/low dimension,) so the qualitative interview data offered the best representation for the complexity of their findings.

Qualitative family research can discover new concepts and terms family members use that crystallize important dimensions of their lives. The CHILD project used child assessment, surveys and questionnaires, and teacher ratings, as well as qualitative fieldwork, observations and interviews in a longitudinal 16-year study of 100 Los Angeles area families with children with generalized developmental delays of various kinds. Parents described their struggles in their own words, and these interviews led to a number of useful constructs that distinguished family accommodation patterns (Gallimore et al. 1996). Families in the study face a familiar and daunting task. They have to re-balance their family lives to accommodate to their child with disabilities. *Accommodation* refers to the process of deciding what activities to do and which not to do given there is a child with disabilities in the family. Accommodation differs from coping with stressors and adaptation, however. It occurs with all levels of stress and responds to perturbations due in part to the child with disabilities, affecting the normal family daily routine. "Look, let me just tell you what I do all day to keep our family together, and then we can talk about supports and stress scales," one mother commented.

Parents frequently used the everyday term, *hassle*; their child was more or less a hassle for them. This is not a pejorative term in parents' everyday use but rather a practical description of the relative disruption and flow of the functional daily routine of activities due at least in part to the child with disabilities. The Ecocultural Family Interview (EFI), a conversational, qualitative method, asks parents about this process (Weisner 2011a). An outcome measure that emerged from this work is *sustainability* of the family, which refers to the attainment of family goals consistent with the moral direction of their lives, as well as the more pragmatic balancing of resources and time (Weisner et al. 2005). Not only parents could describe these circumstances; many adolescents with disabilities themselves, followed since they were age three or four, could provide a reasonable explanatory model of their own illness and sense of difference; "I speak a different dialect from other people," is how one boy described this (Daley and Weisner 2003).

The EFI narratives were summarized and systematically rated along a series of dimensions informed by what parents described to us and by theory from family ecology and research on disability. These quantitative ratings, derived from qualitative interviews and home visits were used along with quantitative family assessment scales to predict child and family outcomes. EFI-derived ratings added significant predictive ability compared to quantitative family assessment scales alone (Nihira et al. 1994).

Qualitative and Fieldwork Methods in Cross-Cultural Family Studies

Understanding world variations in family norms and beliefs. Qualitative fieldwork is essential for including cross-cultural, international, and minority and ethnic communities in family research. Qualitative field studies of families allow researches to understand the remarkably diverse and pluralistic goals, values, beliefs, scripts for everyday activities, and family norms around the world. It is still a useful question to ask: What is arguably the single most important thing to know about a family? The answer, according to some, is exactly where that family lives on Earth, or relatedly, the cultural community in which the family exists. As Therborn (2009) comments regarding his review of the seven broad cross-cultural family systems around the world, “The boys and girls of the world enter many different childhoods and depart them through many different doors” (pp. 338). These family system norms include residence patterns, inheritance laws, gender laws and roles in families, marriage practices, generational and age hierarchies, and many others. Only qualitative fieldwork can capture the variability, adherence, reach, emotional significance, enforcement and extent of influence of the diverse family norms found around the world.

A number of literature reviews find support for the importance of qualitative approaches in understanding families in the global context. For example, LeVine (2007) recently reviewed the ethnographic and qualitative evidence for the remarkable range of parenting across world family systems that have been described in the ethnographic record. Similarly, Barlow and Chapin (2010) reviewed qualitative research on the wide range of mothering and who does mothering in families in diverse cultural communities. Twenty-five years ago, Whiting and Edwards (1988) integrated ethnographic field data and quantitative naturalistic observations of children in family context in 14 communities. Whiting & Edwards show the effects of subsistence ecology, family and household composition, maternal workload and other features on children’s social behavior (nurturance, responsibility, sociability, aggression and others). The amount and importance of these various social behaviors differs dramatically across cultures, as does types of work and family ecology, amount of maternal workloads and supports available, and family norms. Responsibility training, for example, seems fraught and difficult in middle class US families, but far less so in other families around the world, where children show strong task and social responsibility early (Ochs and Izquerdo 2009). *The Child* (Shweder et al. 2009) is an encyclopedic compendium that covers a wide range of family concerns related to socialization, parenting, and child development, and includes cross-cultural, qualitative and ethnographic evidence from around the world. Where the very norms, family goals, and ecology of family life differ dramatically, as they can across cultures, qualitative research and understanding is essential prior to using standard quantitative measures. Even further, qualitative family research in other communities can uncover new family practices that were not thought possible at all, essentially unimagined, in one culture yet which not only occur but are common elsewhere.

Normal variation in family forms and practices is far greater than commonly included in research samples today. For example, 96 % of people studied in the top journals in six sub-disciplines of psychology from 2003 through 2007 were from North America, Europe, Australia, and Israel: "… this means that 96 % of psychological samples come from countries with only 12 % of the world's population" (Henrich et al. 2010, p. 63). "A randomly selected American undergraduate is more than 4000 times more likely to be a research participant than is a randomly selected person from outside of the West," (Henrich et al. 2010, p. 65). Contemporary family research samples are not representative of the family forms and family practices to be found around the world, and our knowledge of the world diversity in family life importantly depends on qualitative accounts of these variations.

Qualitative research broadens understanding of family practices in ethnic minority and international settings that have positive outcomes for children and others. Ethnographic fieldwork provides "existence proofs" for previously unknown family practices that challenge assumptions based only on contemporary Euro–American or Western contexts. For example, the majority of infants and young children around the world co-sleep with parents or others, and the associations of various parenting practices and developmental outcomes with co-sleeping appear quite positive (McKenna and McDade 2005; Morelli et al. 1992; Okami et al. 2002). Older siblings routinely and successfully are asked to care for younger siblings and cousins at ages (typically before 13) when it is thought impossible or dangerous according to US law and Western research (Weisner 1997). Gottlieb (2004) describes one example of the importance of socially distributed caretaking of children from the Beng in the Ivory Coast for example, and Seymour (1999) another among Northeast Indian families in Orissa, many of whom live in large joint households. Although there is a strong emphasis on the importance of adult–child play in US families, such activity, including mother–child play, turns out to be quite rare around the world in the past and still today, based on ethnographic evidence from a large sample of the world's societies (Lancy 2007). Although increased maternal literacy through formal schooling has long been known to be associated with increases in maternal and child health and lower fertility in families around the world, the mechanisms producing this association in fact were not known. A very recent cross-cultural study of family and mothering in Mexico, Venezuela, Nepal, and Zambia, in both rural and urban settings, and in girls and boys with varying levels of schooling, was conducted by LeVine, LeVine, Schnell-Anzola, Rowe and Dexter (2012). The researchers blended community ethnographic study, qualitative interviews and narratives of mothers and children, and quantitative studies of literacy, health, fertility, and language use with children. The team found that maternal increases in literacy through formal schooling led to new *communicative socialization processes* in families and institutions which in turn led to fertility declines, gains in health, and increases in well-being around the world. All these studies used qualitative, ethnographic methods, usually integrated with quantitative measures, to first provide a wider cross-cultural sample, and then search for correlates of the various family and parenting practices of interest.

Conclusion

The suite of qualitative, ethnographic, and fieldwork methods provide remarkable evidence about family forms, family relationships, and family experience. Why would some type of qualitative and contextual evidence *not* be included in many family studies, where feasible and relevant? Many researchers do not have training in qualitative methods. It is recommended that researchers without such training collaborate with those who do have it. Qualitative research is often costly and time consuming. However, costs and time can be managed by using subsamples, nested designs, parallel or "ghost" samples, and by applying for funding when qualitative data clearly can add value to empirical evidence and theory. IRBs and other oversight bodies concerned about and responsible for human participant protections in research currently regularly approve qualitative methods in family research. Others have avoided qualitative approaches because of data and analytic challenges. Today, however, various software packages (Nvivo, Atlas, MaxQDA, Ethnograph, Dedoose) have been developed for relatively quick input and analysis of text and in some cases video data, as well as ways to link to quantitative data. Today, there are also increasingly accepted best practices for the conduct of qualitative research, including criteria for interviewing, focus groups, fieldwork observation, sociolinguistic and narrative methods, and others, as well as ways to analyze the data, and link those findings to quantitative data.

This chapter has outlined a number of qualitative methods used in family research. These approaches have yielded rich, new, and varied insights into the lives of families across the globe. Interviews, observations, collection and review of materials, and other creative techniques form a powerful suite of methodological approaches useful across multiple disciplines. The suite of qualitative methods has been central to research that describes family life, informs family and developmental theories, predicts child and family behavior, and enhances the efficacy and scalability of program implementation. Qualitative and quantitative methods have been creatively married in a number of the most cutting-edge and impactful basic and applied studies in family research. Whether as a stand-alone qualitative or ethnographic family study or part of a mixed methods study, new findings important in family research discovered through qualitative methods continue to enrich the field.

Acknowledgements The New Hope Study has been supported by the MacArthur Network on Successful Pathways Through Middle Childhood, NICHD grant R01HD36038-01A1 (Robert Granger, Aletha Huston, Greg Duncan, and Thomas Weisner, co-PI's), the UCLA Fieldwork and Qualitative Data Laboratory in the Center for Culture and Health (Thomas Weisner, PI), and the Next Generation Project (funded by the David and Lucile Packard, William T. Grant, and the John D. and Catherine T. MacArthur Foundation). The New Hope Ethnographic Study (NHES) is part of the evaluation of New Hope, Inc. conducted by the Manpower Demonstration Research Corporation. The CHILD study was supported by Grants # HD19124 and HD11944 from the National Institute of Child Health and Human Development (Ronald Gallimore, Barbara Keogh, Kazuo Nihira and Thomas Weisner, PIs). The La Vida Project was funded with a grant from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, *Family Obligation and Assistance among Adolescents from Mexican Backgrounds*, R01HD057164 (Andrew Fuligni, PI; Thomas Weisner, Nancy Gonzales, Co-PIs). The study also was supported by the Center for Culture & Health, Semel Institute, Department of Psychiatry, UCLA.

References

Barlow, K., & Chapin, B. L. (Eds.) (2010). Special issue: Mothering as everyday practice. *Ethos, 38*(4), 321–322.

Bernard, H. R. (2013). *Social research methods: Qualitative and quantitative approaches* (2nd ed.). Los Angeles: Sage.

Borneman, J., & Hammoudi, A. (Eds.). (2009). *Being there: The fieldwork encounter and the making of truth*. Berkeley: University of California Press.

Bronfenbrenner, U. (1995). Developmental ecology through space and time: A future perspective. In P. Moen, G. H. Elder, K. Luscher, & U. Bronfenbrenner (Eds.), *Examining lives in context: Perspectives on the ecology of human development* (pp. 619–647). Washington, DC: American Psychological Association.

Bronfenbrenner, U. (Ed.). (2005). *Making human beings human: Biocultural perspectives on human development*. Thousand Oaks: Sage Press.

Creswell, J. W., Klassen, A. C., Plano-Clark, V. L., & Clegg Smith, K. (2011). Best practices for mixed methods research in the health sciences. http://obssr.od.nih.gov/scientific_areas/methodology/mixed_methods_research/pdf/Best_Practices_for_Mixed_Methods_Research.pdf.

Cross-Barnet, C., Cherlin, A., & Burton, L. (2011). Bound by children: Intermittent cohabitation and living together apart. *Family Relations, 60,* 633–647.

Daley, T., & Weisner, T. S. (2003). "I speak a different dialect": Teen explanatory models of difference and disability. *Medical Anthropology Quarterly, 17*(1), 25–48.

D'Andrade, R., & Strauss, C. (1992). *Human motives and cultural models*. Cambridge: Cambridge University Press.

Dedoose (Version 4.3.87) [Web application for managing, analyzing, and presenting qualitative and mixed method data]. (2013) Los Angeles, CA: SocioCultural Research Consultants, LLC.

Duncan, G., Huston, A., & Weisner, T. S. (2007). *Higher ground: New Hope for the working poor and their children*. New York: Russell Sage Foundation.

Gallimore, R., Coots, J. J., Weisner, T. S., Garnier, H., & Guthrie, G. (1996). Family responses to children with early developmental delays II: Accommodation intensity and activity in early and middle childhood. *American Journal of Mental Retardation, 101*(3), 215–232.

Gibson, C., & Weisner, T. S. (2002). "Rational" and ecocultural circumstances of program take-up among low-income working parents. *Human Organization, 61*(2), 154–166.

Gilgun, J. (2012). Enduring themes of qualitative family research. *Journal of Family Theory & Review, 4,* 80–95.

Gottlieb, A. (2004). *The afterlife is where we come from. The culture of infancy in West Africa*. Chicago: University of Chicago Press.

Greene, J. C. (2007). *Mixed methods in social inquiry*. San Francisco: Jossey-Bass.

Harkness, S., Zylicz, P. O., Super, C. M., Welles-Nyström, B., Bermúdez, M. R., Bonichini, S., ... Mavridis, C. J. (2011). Children's activities and their meanings for parents: A mixed-methods study in six Western cultures. *Journal of Family Psychology, 25*(6), 799–813.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33,* 61–135.

Kohn, M. L. (1977). *Class and conformity: A study of values*, (2nd ed.). Chicago: The University of Chicago Press.

Lancy, D. (2007). Accounting for variability in mother–child play. *American Anthropologist, 109*(2), 273–284.

Lareau, A. (2003). *Unequal childhoods: Class, race, and family life*. Berkeley: University of California Press.

LaRossa, R. (2005). Grounded theory methods and qualitative family research. *Journal of Marriage and Family, 67,* 837–857.

LeVine, R. A. (2007). Ethnographic studies of childhood: A historical overview. *American Anthropologist, 109,* 247–260.

LeVine, R. A., LeVine, S., Schnell-Anzola, B., Rowe, M. L., & Dexter, E. (Eds.). (2012). *Literacy and mothering: How women's schooling changes the lives of the world's children*. New York: Oxford University Press.

Lieber, E. (2009). Mixing qualitative and quantitative methods: Insights into design and analysis issues. *Journal of Ethnographic and Qualitative Research, 3,* 218–227.

McKenna, J. J., & McDade, T. (2005). Why babies should never sleep alone: A review of the co-sleeping controversy in relation to SIDS, bedsharing and breast feeding. *Paediatric Respiratory Reviews, 6*(2), 134–152.

Morelli, G. A., Rogoff, B., Oppenheim, D., & Goldsmith, D. (1992). Cultural variations in infants' sleeping arrangements: Questions of independence. *Developmental Psychology, 28*(4), 604–613.

Nihira, K., Weisner, T., & Bernheimer, L. (1994). Ecocultural assessment in families of children with developmental delays: Construct and concurrent validities. *American Journal of Mental Retardation, 98*(5), 551–566.

Ochs, E., & Izquerdo, C. (2009). Responsibility in childhood: Three developmental trajectories. *Ethos, 37*(4), 391–413.

Okami, P., Weisner, T. S., & Olmstead, R. (2002). Outcome correlates of parent-child bedsharing: An eighteen year longitudinal study. *Journal of Developmental and Behavioral Pediatrics, 23*(4), 244–253.

Pearce, L. D., & Denton, M. L. (2011). *A faith of their own: Stability and change in the religiosity of America's adolescents*. New York: Oxford University Press.

Roy, K., & Burton, L. (2007). Mothering through recruitment: Kinscription of nonresidential fathers and father figures in low-income families. *Family Relations, 56,* 24–39.

Roy, K., Buckmiller, N., & McDowell, A. (2008). Together but not "together": Trajectories of relationship suspension for low-income unmarried parents. *Family Relations, 57,* 198–210.

Seymour, S. C. (1999). *Women, family, and child care in India: A world in transition*. Cambridge: Cambridge University Press.

Shweder, R., Bidell, T., Dailey, A., Dixon, S., Miller, P., & Modell, J. (Eds.). (2009). *The child: An encyclopedic companion*. Chicago, IL: University of Chicago Press.

Small, M. L. (2011). How to conduct a mixed methods study: Recent trends in a rapidly growing literature. *Annual Review of Sociology, 37,* 57–86.

Steinberg, M. (2012). *La Vida: Negotiating family responsibility and obligation among Mexican adolescents in Los Angeles*. Poster session presented at the meeting of Society for Research in Child Development (SRCD) Themed Meeting: Positive Development of Minority Children, Tampa, Florida. (2012, February).

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks: Sage.

Therborn, G. (2009). Family. In R. A. Shweder, T. R. Bidell, A. C. Dailey, S. Dixon, P. J. Miller, & J. Modell (Eds.), *The child: An encyclopedic companion* (pp. 333–338). Chicago: University of Chicago Press.

Webb, E. T., Campbell, D. T., Schwartz, R. D., Sechrest, L., & Grove, J. B. (1981). *Nonreactive measures in the social sciences, Second Edition*. Boston: Houghton Mifflin.

Weisner, T. S. (1996). Why ethnography should be the most important method in the study of human development. In R. Jessor, A. Colby, & R. Shweder (Eds.), *Ethnography and human development. Context and meaning in social inquiry* (pp. 305–324). Chicago: University of Chicago Press.

Weisner, T. S. (1997). The 5 to 7 transition as an ecocultural project. In A. Sameroff & M. Haith (Eds.), *The five to seven year shift: The age of reason and responsibility* (pp. 295–326). Chicago: University of Chicago Press.

Weisner, T. S. (Ed.). (2005). *Discovering successful pathways in children's development: Mixed methods in the study of childhood and family life*. Chicago: University of Chicago Press.

Weisner, T. S. (2011a). The Ecocultural Family Interview: New conceptualizations and uses for the study of illness. In S. Bonichini & M. R. Baroni (Eds.), *Papers in honor of Vanna Axia* (pp. 166–180). Padova: University of Padova Press.

Weisner, T. S. (2011b). If you work in this country you should not be poor, and your kids should be doing better: Bringing mixed methods and theory in psychological anthropology to improve research in policy and practice. *Ethos, 39*(4), 455–476.

Weisner, T. S., & Duncan, G. (in press). The world isn't linear or additive or decontextualized: Pluralism and mixed methods in understanding the effects of anti-poverty programs on children and parenting. In E. T. Gershoff, R. S. Mistry, & D. A. Crosby (Eds.), *Societal contexts of child development: Pathways of influence and implications for practice and policy*. New York: Oxford University Press.

Weisner, T. S., & Fiese, B. H. (2011). Introduction to the special section of the *Journal of Family Psychology*: Integrative and applied solutions for family science. *Journal of Family Psychology*, *25*(6), 795–798.

Weisner, T. S., Matheson, C., Coots, J., & Bernheimer, L. P. (2005). Sustainability of daily routines as a family outcome. In A. E. Maynard & M. I. Martini (Eds.), *Learning in cultural context: Family, peers and school* (pp. 47–74). New York: Kluwer/Plenum.

Whiting, B. B., & Edwards, C. P. (1988). *Children of different worlds: The formation of social behavior*. Cambridge: Harvard University Press.

Yoshikawa, H., Weisner, T. S., & Lowe, E. (Eds.). (2006). *Making it work: Low-wage employment, family life, and child development*. New York: Russell Sage Foundation.

Yoshikawa, H., Weisner, T. S., Kalil, A., & Way, N. (2008). Mixing qualitative and quantitative research in developmental science: Use and methodological choices. *Developmental Psychology, 44,* 344–354.

Chapter 11
Approaches to Measuring Families

Paul Amato

Measurement is concerned with what can be observed, the conditions under which observations are made, and how observations are recorded for future analysis and consideration. Family researchers have used a variety of measurement approaches, including the direct observation of family behavior, surveys, and interviews. In recent years, three emerging methods—ecological momentary assessment (EMA), biomarkers, and mixed qualitative-quantitative designs—have captured the interest of family scholars. In the first part of this chapter, I comment on each of these measurement approaches. In the second part of this chapter, I address a broader issue: how researchers can measure the characteristics of entire families rather than individual family members. My comments are from the perspective of a researcher who is interested in measurement issues but is not primarily a methodologist.

Ecological Momentary Assessment

Smyth and Heron (Chap. 9) provide an informative description of ecological momentary assessment (EMA)—a recently developed methodology with the potential to provide family scholars with new insights into the links between family processes and people's thoughts and moods. A major advantage of this method is that it collects data from people in everyday, natural settings.

Much of what we know about families has been collected through surveys of various types. Although survey research has many advantages, it is low in ecological validity. In a typical survey, people respond to questions while speaking on the telephone, participating in face-to-face interviews, completing questionnaires, or interacting with computers. Of necessity, surveys require people to step outside of the normal flow of family events and interactions. Studies in which family members

P. Amato (✉)
Department of Sociology and Crime, Law, & Justice,
The Pennsylvania State University, University Park, PA 16802, USA
e-mail: pamato@la.psu.edu

S. M. McHale et al. (eds.), *Emerging Methods in Family Research,*
National Symposium on Family Issues 4, DOI 10.1007/978-3-319-01562-0_11,

are observed in laboratory settings are even less natural. EMA, in contrast, collects information from people while they are in participating in everyday activities. Because people report on their *current* experiences, EMA avoids the biases inherent in recalling events. And because multiple measurements are taken, it is possible to model the temporal dynamics of people's experiences. Given these advantages, EMA is likely to become a valuable supplement to more conventional survey and observational approaches.

Although EMA is useful for studying everyday activities, researchers would find it difficult to study extraordinary events or unusual settings with this method. For example, it's not clear how one would study topics like crime victimization, violence, or emergency intervention. In contrast, people spend a great deal of time at home with their families engaging in routine activities, such as eating meals, working around the house, and participating in leisure activities. If a typical person is "paged" to record information at random times during the day, many of these episodes will occur when the person is at home with family members. For this reason, EMA is well suited to capture people's experiences as they participate in the general ebb and flow of family life.

EMA provides rich information about the moods, thoughts, and behaviors of *individuals*. This method can be adapted to provide information about *families* by including assessments from multiple family members. Indeed, EMA is probably the best available method for studying how the experiences of family members covary over time. Smyth and Heron (Chap. 9) refer to several studies that illustrate this potential. For example, Slatcher, Robles, Repetti, and Fellows (2010) found that husbands' reports of work stress (collected through EMA) were associated with elevated cortisol levels among wives. The association was weaker, however, when wives were high in marital satisfaction, which suggests a buffering effect of positive marital quality. Husbands' cortisol levels, in contrast, were not associated with wives' work stress.

Although EMA's potential is substantial, only a small number of studies have used it to investigate family life. The following are a few examples of research topics and questions that could be studied with multi-person EMA designs.

1. With respect to the transmission of mood within the family, it should be possible, with multiple assessments and cross-lagged models, to assess the direction of influence between family members. For example, does mood transmission run mainly from husbands to wives or from wives to husbands? Similarly, does mood transmission run primarily from parents to children or from children to parents? Or is influence generally reciprocal?
2. Are parents' reports of marital conflict on particular days associated with increases in children's reports of negative affect? Correspondingly, are children's reports of negative affect associated with increases in tension between parents?
3. On days when husbands perform a significant amount of housework, do their mood and relationship happiness ratings decline? On the same days, do wives' ratings of mood and relationship happiness increase? Is this pattern reversed on days when husbands do no housework and wives do a great deal? Are these

patterns moderated by spouses' hours of paid employment and traditional attitudes about gender roles in marriage?

4. How well do spouses agree on particular marital events, such as arguments? Do husbands tend to rate the seriousness of arguments differently than wives do? How might discrepancies vary with the topic of the argument or the person who raises the issue? Does perceived seriousness affect the length of time it takes spouses to recover with respect to mood and feelings for the partner?

Biomarkers

Halpern and Harris (Chap. 8) begin with the reasonable assumption that human behavior reflects influences at multiple levels, including biological influences from "below" and social, cultural, and environmental influences from "above." Consequently, to obtain the fullest understanding of behavior, social scientists need to add biological variables to their theories and data collection efforts. Halpern and Harris draw from their experiences with the National Longitudinal Study of Adolescent Health (Add Health) to provide useful information on different types of biomarkers, how to incorporate biomarkers into survey research, and the potential of biomarkers to complement survey data.

Collecting biospecimens in the field is complicated, especially when a study is national in scope. Researchers must obtain the cooperation of respondents; train field-workers in methods of extracting, handling, and transporting biological specimens; ensure that labs have adequate capacity; and select the most appropriate laboratory methods in a field that is evolving rapidly. The list of things that can go wrong is daunting, and the authors' description of the problems they encountered in obtaining saliva samples is instructive. But despite these challenges, the ongoing collection of biomarkers in Add Health (and in other studies) indicates that it is possible to collect biological data successfully in the context of large-scale national surveys.

The inclusion of biological data into models of social behavior is a positive development in the social sciences. To paraphrase Edward O. Wilson (2012), to understand society, we must understand history. To understand history, we must understand prehistory. And to understand prehistory, we must understand biology. A continuous line connects the present with the past, and there is no reason to assume that biological influence withered away at some point. Instead, genetic and cultural evolution occurred alongside one another over long periods of time to produce what we think of as human nature.

Although the goal of a truly integrated biosocial science is far in the future, Halpern and Harris (Chap. 8) provide several examples of how contemporary researchers are combining biological and survey data in innovative ways. Bruckner and Bearman (2005) showed that youth who made virginity pledges did not differ from other youth in their infection rates from sexually transmitted diseases (based on urine samples obtained in the Add Health study). And Miller et al. (2011) found that children raised in families with low socioeconomic status had elevated rates of chronic illness in adulthood (based on biological health data collected as part of the Midlife in the United States study). The association between childhood socioeconomic status and illness was weaker, however, among adults who recalled a high level of maternal nurturance. Studies like these demonstrate how biological data can be combined with survey data to study social influences on health outcomes.

As a family scholar, I find current research on genotype-environment interactions to be of particular interest. For example, it is well known that stressful family circumstances, such as parental divorce or poverty, increase the risk of a variety of problems for children. Yet many children from troubled families develop into happy, competent, well-adjusted young adults. For decades researchers have tried to understand why some children are vulnerable to stress and others are resilient. Several recent studies suggest that children's susceptibility has a genetic origin (Belsky and Pluess 2009; Guo et al. 2008; Simons et al. 2011). These studies have found that alleles (variations) of particular genes appear to increase or decrease the probability that stressful family conditions are linked with problematic outcomes among children, such as delinquency, depression, and academic failure.

A recent study along these lines by Simons and colleagues (2011) is a good example of this line of research. They examined polymorphisms in 5-HTTLPR and DRD4—genes that affect levels of serotonin and dopamine in the central nervous system, respectively. Some observers have proposed that these genes regulate people's thresholds for pleasure and pain and, hence, their responsiveness to environmental rewards and punishments (Belsky and Pluess 2009). Simons and his colleagues considered the role of these genes in moderating the association between a measure of social adversity (which included harsh parenting, neighborhood victimization, and discrimination) and a measure of youth aggression (which included damaging property and fighting with weapons). Among people with a genotype suggesting low plasticity, social adversity was not associated with aggression. But among people with a genotype suggesting high plasticity, social adversity was positively associated with aggression. Importantly, among individuals with high genetic plasticity, a high degree of adversity was associated with an above-average level of aggression, whereas a low degree of adversity was associated with a below-average level of aggression. In other words, individuals with a genotype linked to plasticity appeared to suffer from a high level of adversity but to benefit from a low level of adversity.

Given that the search for gene-family environment interactions is relatively recent, it is not clear how productive this line of work will be. New and innovative findings often turn out to be false positives. Nevertheless, these studies appear to have a great deal of potential to help us understand how children—and perhaps adults— respond to their social and physical environments.

Qualitative and Mixed-Method Approaches

The chapter on qualitative methods by Weisner (Chap. 10) might seem out of place in a discussion of measurement. Some textbooks define measurement as a systematic procedure for assigning numbers to observations (e.g., Knoke and Bohrnstadt 1991,

p. 9). If one accepts this definition, then qualitative approaches are beside the point. But if one thinks of measurement more broadly, that is, as a systematic procedure for recording observations for analysis and further consideration (e.g., Babbie 1995, p. 2), then qualitative as well as quantitative methods involve "measurement."

Quantitative and qualitative methods both aim to describe the world, but in one case observations are recorded as numbers (or later transformed into numbers), and in the other case observations are recorded as words (or visual images). People think in terms of numbers (How much? How often?), words, and visual images, and each represents a different way of understanding reality. Indeed, different parts of the brain appear to specialize in numeric, verbal, and spatial reasoning (Just and Varma 2007). If this is true, then shouldn't researchers use all of their brains to study social phenomena? Researchers translate words (or images) into numbers to manipulate the data mathematically. Researchers also translate numeric results into words and images, not only to communicate with others, but also to provide a deeper understanding of the phenomenon. Shifting from one data mode to another allows us to use all of the tools in our intellectual repertoire.

Although qualitative methods have been around for a long time, a more recent trend has been to combine qualitative and quantitative methods in the same research project. Several mixed-method studies in which qualitative and quantitative data nicely complement one another are described by Weisner (Chap. 10). In the New Hope study, for example, qualitative interviews revealed why some participants used program components less frequently than the experimenters had anticipated. Qualitative data also illuminated why some parents were more successful than others in obtaining good jobs and increasing their income. In some cases, the qualitative data suggested new directions for those working on the quantitative side. As these examples illustrate, qualitative and quantitative approaches illuminate different facets of the social world, and they have much to offer when combined in the same study.

Although we sometimes think of quantification as being synonymous with science, it is useful to recall that qualitative approaches play an important role in all the sciences, including the natural sciences. Charles Darwin's field research and his theory of evolution through natural selection is a good example. Darwin spent five years traveling around the world on the Beagle, collecting observations of animals, insects, plants, fossils, and local geology, and keeping copious notes that spanned many volumes. During this time Darwin came to accept the "transmutation of species," which paved the way for his later ideas about evolution. Darwin's *On the Origin of Species* (1859)— one of the greatest scientific books every written—is remarkable in that it contained almost no math. To make his case, Darwin described the variation within and between species, the geographical distribution of species, the fossil record, and the results of animal breeding—all of which can be thought of as qualitative evidence.

Ethology (the study of animal behavior) is a science that relies substantially on qualitative observations. Ethologists' field notes take two general forms: an ethogram—a qualitative list of the behaviors observed, and a time budget—a quantitative summary of the percentage of time animals spend in various behaviors (Lerner

1996). Ethograms are frequently supplemented with visual aids, such as drawings, photographs, and videos. Some of the most remarkable findings from this field have emerged from direct (unquantified) observation. Jane Goodall—undoubtedly the world's most famous ethologist—spent years observing chimpanzee behavior at Gombe, Tanzania. Her research revealed that chimpanzees make and use tools, and that young chimps learn this behavior from observing older members of the troop (Goodall 1988). Because culture consists of behavior that is "learned and shared," chimpanzees possess something akin to human culture—a remarkable revelation. Although Goodall was not trained in quantitative methods (and in fact had little training of any kind when she first went into the field), she was an excellent observer. Indeed, biologist Stephen Jay Gould referred to her work as nothing less than "one of the Western world's great scientific achievements" (Gould 1988, p. v). Like most scientists, ethologists also rely on quantitative models, tests of statistical significance, and experimentation. But most ethologists accept that a qualitative, descriptive phase is a necessary prerequisite and complement to numerical analysis (Hinde 1987).

At the time of this writing, important qualitative data are beaming to earth from the Mars rover, Curiosity. In August of 2012, Curiosity sent pictures of an ancient streambed located on the Martian surface. For decades, scientists had hypothesized the existence of streambeds on Mars, but this was the first time anyone had ever seen one. Scientists realized they were viewing a streambed without making use of Curiosity's sophisticated onboard tools for chemical analysis. The key evidence was the rounded shape of gravel, comparable to gravel on earth that has been bounced around for long periods of time by moving water (Landau 2012). In other words, scientists reached their conclusion through a straightforward process of pattern recognition. They know what streambeds on earth look like; Curiosity's images *look* like a streambed, so scientists concluded that it is a streambed. Despite the great value of quantification and mathematical analysis, sometimes great discoveries in the natural sciences are based on unaided, direct observation.

Measuring Family Characteristics

How do researchers measure characteristics of families, as opposed to the characteristics of individual family members? In addressing this question, it is useful to distinguish between aggregate, structural, and relational family characteristics. Researchers measure aggregate characteristics by adding up (or taking the mean of) individual characteristics. For example, total family income can be measured as the sum of the earnings of each family member in a household. With respect to a married couple, the mean age of the husband and wife would be an example of an aggregate characteristic.

Structural characteristics refer to who is in the family (with respect to number, gender, and generation), the basis of family membership (blood, marriage, or adoption), and people's roles in the family (i.e., spouse, parent, child, grandparent, stepparent). Combining these elements produces the typical demographic categories used in family research: two-parent families with children, single-mother families, three generation extended households, childless married couples, same-sex couples with children, and so on. (As I note later, however, structural characteristics are better viewed as features of households rather than families.)

Relational characteristics are based on the relationships between family members. Examples of relational characteristics include the amount of conflict between family members, the existence of cliques or alliances within the family, and whether decision-making power is autocratic or democratic. To measure relational characteristics, researchers must obtain information on how family members feel about and interact with one another.

Attention to relational characteristics follows naturally from a system's perspective. But despite the widespread recognition that families are systems, relatively little research has focused on the direct measurement of family characteristics. Qualitative research may be an exception. As Weisner (Chap. 10) points out, qualitative researchers often pay attention to the larger systems in which people are embedded. Through participant observation, researchers learn how groups are organized, what the group norms are, what roles people play, and how group members routinely interact with one another. Quantitative researchers, in contrast, have tended to focus on individual attributes, such as marital satisfaction, parent's child-rearing behaviors, or health outcomes—presumably because of the abundance of large-scale survey data sets.

Despite the individualistic focus of most quantitative work, a few quantitative researchers have made sustained efforts to measure relational characteristics of families. This work, however, is scattered across multiple research literatures. To illustrate these efforts, I describe three examples, one from the field of social psychology and two from family therapy.

A Social Psychological Approach

Some social psychologists have attempted to measure the attributes of primary groups, that is, groups in which members have frequent face-to-face interaction. In the 1970s and 1980s, William and Ruth Scott (1981) initiated a program of research that focused on identifying and measuring the fundamental dimensions of small groups, including families. (Although they referred to these dimensions as "structural," they are relational in the sense I use in this chapter.) They focused on six dimensions, which were derived from the classic works of social theorists such as Karl Marx and Max Weber.

1. *Permeability* is the ease of entering or leaving the group.
2. *Consensus* is the degree to which members share norms, attitudes, and beliefs.
3. *Solidarity* refers to cohesiveness and a preference for interaction within the group.
4. *Fragmentation* refers to the existence of cliques within the group.
5. *Role differentiation* involves a clear division of labor with distinctive roles.
6. *Status differentiation* refers to differences in power and status.

The Scotts measured these constructs with a combination of sociometric indices, insider (family member) reports, and outsider reports. They also believed that it was necessary to obtain information from *all* group members to measure group constructs adequately. Scott and Scott (1981) found a variety of significant associations between these dimensions of families. For example, families with high levels of agreement between members tended to be cohesive and have a relaxed division of labor. Cohesive families also had relatively little turnover in membership. Moreover, families with a high level of inequality tended to be split into cliques and subgroups.

The Circumplex Model

David Olson's (2000) Circumplex Model was developed specifically for the clinical assessment of families. This model focuses on three dimensions of family life that, according to Olson, have emerged repeatedly in the family therapy literature.

1. *Cohesion* is the emotional bonding between family members. It ranges from disengaged (very low) to separated (low) to connected (high) to enmeshed (very high).
2. *Flexibility* refers to the amount of change in leadership, role relationships, and relationship rules. It ranges from rigid (very low) to structured (low) to flexible (high) to chaotic (very high).
3. *Communication* involves responsiveness, clarity, self-disclosure, and speaking/listening skills.

The model assumes that extreme values of cohesion or flexibility (high or low) are problematic, and that well-functioning (balanced) families lie in the midrange of these two dimensions. Families presenting for treatment are usually unbalanced in some fashion: disengaged, enmeshed, rigid, or chaotic. Accordingly, the general goal of family therapy is to move families from an unbalanced to a balanced system. Communication, the third dimension, is a facilitating factor that helps to move families along the first two dimensions.

Olson developed the Family Adaptability and Cohesion Evaluation Scales (FACES) to assess cohesion and flexibility (adaptability). These instruments involve family members' self-reports or therapists' ratings, and scores on these two dimensions allow families to be classified as balanced or unbalanced types. Research with these instruments has shown that the families of schizophrenics, neurotics, and sex offenders are more likely to be unbalanced than are the families of individuals not in therapy (Olson 2000). People from unbalanced families also report more family stress and less satisfaction with family life (Olson 2011).

The Family Environment Scale

The Family Environment Scale (FES) also emerged from the clinical field (Moos and Moos 1994). Family counselors and therapists use this instrument to identify issues in family treatment and to monitor improvement over time. The various dimensions of the scale were based on the authors' observations and interviews with families. Three versions of the instrument allow people to evaluate their families (a) as they are (b) as they would be in a perfect world, and (c) as they are expected to be in the future. The instrument has been used to assess the family environment from the perspective of a single member or from the perspectives of different members of the same family.

The 10 scales on the FES are grouped into three areas: family relationships (3 scales), system maintenance (2 scales), and personal growth (5 scales).

1. *Cohesion* is the degree of commitment and support family members provide to one another.
2. *Expressiveness* is the extent to which family members are encouraged to express their feelings.
3. *Conflict* is the amount of openly expressed anger and conflict among members.
4. *Organization* refers to the extent to which family activities are planned.
5. *Control* is the extent to which rules and procedures guide family life.
6. *Independence* is the extent to which family members are self-sufficient and make their own decisions.
7. *Achievement* orientation is the extent to which family activities are competitive and oriented toward success.
8. *Intellectual-cultural orientation* reflects family members' interest in political, intellectual, and cultural activities.
9. Active-recreational orientation refers to family members' participation in social and recreational activities.
10. *Moral-religious orientation* is the extent to which the family emphasizes ethical and religious issues and values.

Although the FES originally was developed for clinical research, some researchers have used this instrument with nonclinical samples. For example, one study found that mothers' ratings of family cohesion on the FES were negatively associated with teachers' ratings of children's behavior problems (Lucia and Breslau 2006). Another study found that ratings on the cohesion, active-recreational, and intellectual-cultural dimensions of the FES were positively associated with friendship quality among college women (Wise and King 2007).

Limitations of Previous Approaches

These examples are typical of existing approaches to measure relational family characteristics. Although these contributions are useful, several problems are inherent in this work. First, on a conceptual level, it is not clear how to identify the key dimensions of family life. Researchers have identified these dimensions either through reading classic social theory (Scott and Scott 1981) or relying on clinical experience and insight (Moos and Moos 1994; Olson 2000). But in the absence of an explicit theory of family relationships and functioning, the list of currently studied dimensions is, to a certain extent, arbitrary.

A second problem involves measurement strategies. Most attempts to measure family relational characteristics have relied on the ratings of insiders (family members) or outsiders (therapists or trained observers). But it is not clear what one should do when the ratings of reporters differ. For example, it would not be surprising to find that parents and adolescents in the same family tend to have different views about topics such as authority and conflict. Indeed, parents tend to see family relations as being fairer and more peaceful than do adolescents (Smetana 1988). In cases like these, does it make sense to average the ratings of different observers? Or should disagreements about family relations be incorporated somehow into the measurement? One could imagine, for example, generating the mean level of reported conflict for a family (based on the ratings of all family members) as well as the variance of these ratings (to assess disagreement between raters).

A third problem involves how to define the unit to be measured. How does a researcher determine who is in a family? This question is usually answered in an objective fashion, that is, researchers "impose" their definitions of family on the people being studied. Anthropologists call this an *etic* perspective. For example, the U.S. Census Bureau (2012) defines a family in the following manner: "A family is a group of two or more people related by birth, marriage or adoption and residing together." This definition is useful for obtaining a current snapshot of various family configurations (for example, the percentage of children living with single parents in a given year), and to track changes over time in family living arrangements. The limitation of this definition is that most people do not define families in this fashion.

An alternative (*emic*) approach is to let people make their own decisions about who is in their families. Consider the hypothetical people displayed in the upper left portion of Fig. 11.1. Three individuals—a mother, the mother's child, and the stepfather of the child (the mother's second husband)—live in a household, which is indicated by a rectangle. A researcher using the U.S. Census Bureau definition would treat these three people as a family. A researcher with a different perspective, however, might ask each individual in the household, "Who is in your family?" (See Levin 1993; Widmer 1999, 2006 for examples of this approach).

The mother might list four people: her child, her husband, her own mother (the child's grandmother), and her best friend from high school (with whom she has maintained close contact over the years). The stepfather lists his wife and stepchild, but does not list his mother-in-law, with whom he is not close. He lists the two children from his first marriage, however, who live in a different household. Finally, the child lists his grandmother and mother but does not list his stepfather. Indeed, many children do not view their stepparents as family members (Furstenberg 1987). Nevertheless, the child includes his father and his father's new wife, with whom he gets along well (possibly because the child and the stepmother do *not* live together).

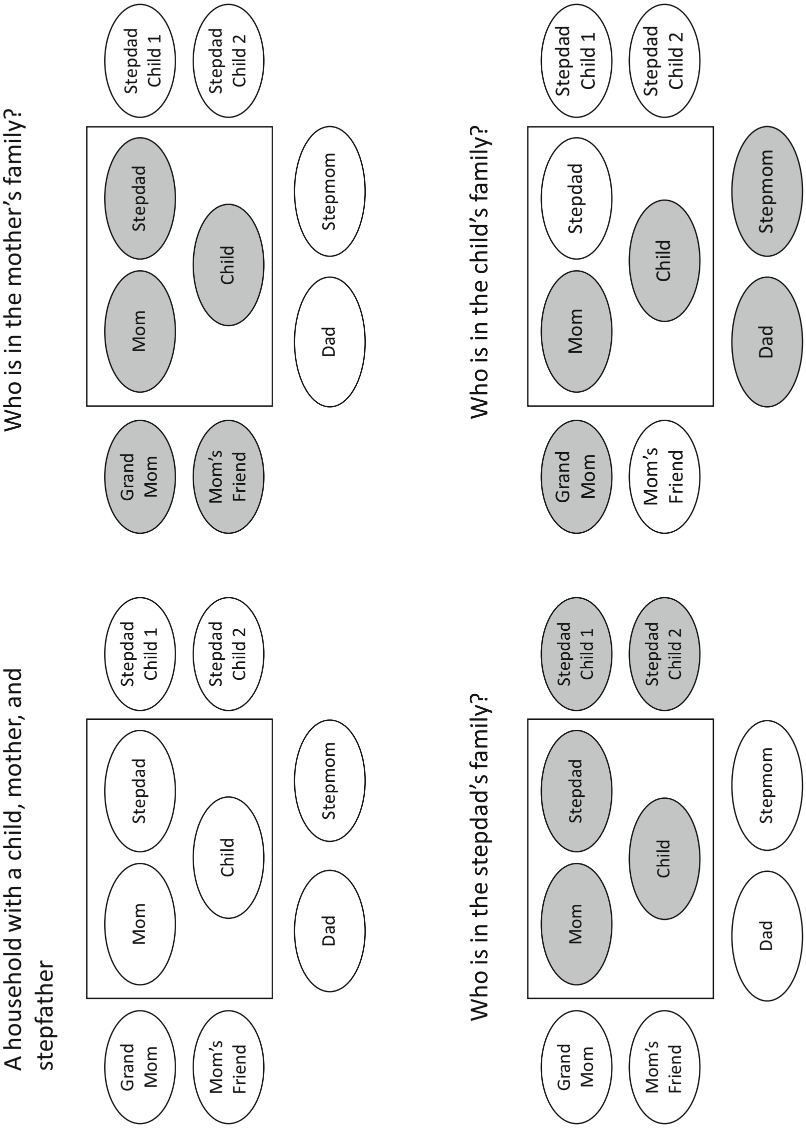


Fig. 11.1 Hypothetical networks of household members

Defining families subjectively yields several conclusions that diverge sharply from the more typical approach of defining families objectively and equating them with households. First, as the mother's choices in Fig. 11.1 indicate, not all family members are related by blood, marriage, or adoption. Indeed, considering close friends to be family members appears to be quite common (Widmer 2006). Many social scientists have noted the importance of fictive kin in people's lives—a phenomenon that may be especially important for some racial and ethnic groups, such as African Americans (Chatters et al. 1994). From a subjective perspective, of course, these individuals are not "fictive."

Second, families are spread across more than one household. This is not a new insight; many observers have pointed out that families are not the same as households (e.g., Cherlin & Furstenberg 1994; Levin 1993; Widmer 1999). But family researchers have rarely addressed the implications of this fact for measurement. Moreover, one can argue that the tendency to equate families with households has led researchers to give insufficient attention to relationships and interactions with extended family members (Sarkisian and Gerstel 2012).

Third, family membership is asymmetrical. That is, person A might claim person B as a member of his/her family, but person B might not claim person A (i.e., the child and stepfather in Fig. 11.1). In contrast, when families are defined objectively, membership is always reciprocal. And fourth, people living in the same household can belong to different families. This is true for each of the three individuals in Fig. 11.1.

A Social Network Approach

Taking the respondent's perspective seriously creates challenges for data collection and analysis. Nevertheless, social network methods can be adapted to study relational characteristics based on people's subjective definitions. Researchers can identify family networks, as noted earlier, by asking focal respondents to list all the members of their families. Researchers might provide verbal aids to assist people in this task. For example, respondents could be instructed to "list any family members who are important to you or play a significant role in your life." A common result of this approach would be that members of the same household list different family networks, as suggested earlier. But that would not stop researchers from inquiring about the perceived characteristics of these family networks. For example, respondents could report on the extent to which family members argue or disagree with one another, despite the fact that each person in a household might be referring to a somewhat different group of people. Although this approach might seem unwieldy, it reflects the reality of many people's lives these days.

Other characteristics that may be particularly applicable to family networks include the extent to which members are emotionally close, are in contact with one another, exchange assistance, or are available as potential help givers in times of need. Obtaining this information would allow researchers to characterize family networks as cohesive versus fragmented, for example. A safe assumption (which could be tested) is that children with divorced parents report more fragmented family networks than do children with continuously married parents. This type of information would be useful in allowing researchers to assess the extent to which family networks are sources of social capital. Putnam (2000) described social capital as the total value of a person's social networks with respect to cooperation and mutual support. Given that virtually all theorists have viewed the family as one of the main sources of social capital (e.g., Bourdieu 1993; Coleman 1988; Putnam 2000), this construct is directly relevant to the study of family networks. Presumably, people with large and cohesive networks are advantaged in a variety of ways, although networks also can be a source of strain.

In addition to assessing the characteristics of people's family networks, researchers can assess the links between overlapping networks. In Fig. 11.1, for example, the child and the mother claim each other as family members. It might be of interest, therefore, to assess the extent to which these two networks overlap. In this case, the mother's family network contains five members (including herself), and three of these individuals also are members of the child's family network. Correspondingly, three of the five people in the child's family network are also members of the mother's family network. Consequently, we can say that these two networks overlap by 60 %. By extension, one might create an index of network overlap for all three members of the household: the mother, the stepfather, and the child. Researchers could then relate this figure to various individual-level outcomes—perhaps the mother's marital satisfaction or the child's educational achievement. The assumption underlying the analysis would be that households with a high level of family network overlap confer certain benefits on their members. Or one might argue for the alternative hypothesis when these networks have a great deal of tension.

In addition to assessing links between overlapping networks, researchers could assess the extent to which members of one family network have relations with or exchange resources with members of different networks. For example, in Fig. 11.1, one might inquire about the links between the mother's best friend and the stepfather's children, or between the grandmother and the father's wife (the stepmother). What are the implications of the fact that these people are not in the same family network but can be connected through other network members?

These latter questions are related to the distinction, often made by network researchers, between *bonding* social capital and *bridging* social capital (Putnam 2000). Bonding social capital refers to the degree of belongingness and reciprocity in a person's network, whereas bridging social capital refers to overlapping networks through which individuals might receive some benefit. That is, a member of one network might gain resources from a different network because the two networks have one or more people in common (i.e., a friend of a friend). Numerous scholars have studied families and social capital (e.g., Furstenberg 2005; Ravanera and Rajulton 2010), especially the implications of family social capital for children's well-being (e.g., Sandefur et al. 2006; Crosnoe 2004). In general, however, family scholars working in this area have not viewed the definition of family as problematic and have not considered how people's subjective definitions of family might affect social capital. (For an exception, see Widmer 2006).

In conclusion, to study families as *families* and not as aggregates of individuals, researchers must define the unit of analysis. Although an etic approach (in which the researcher's definitions are used) is necessary to answer many research questions, this approach clashes with people's everyday understandings of families. The notion that families extend beyond households is widely recognized, but previous research has not been clear about how to study families as distinct from households. Viewing families as overlapping networks that extend across multiple households, with each network having at its nucleus a reference person, might yield new insights, especially in an era when families are becoming more complex and difficult to classify.

References

Babbie, E. (1995). *The practice of social research* (7th ed.). Belmost, CA: Wadsworth.

Belsky, J., & Pluess, M. (2009). Beyond diathesis stress: Differential susceptibility to environmental influences. *Psychological Bulletin, 135,* 885–908.

Bourdieu, P. (1993). *Sociology in question*. London: Sage.

Brückner, H., & Bearman, P. (2005). After the promise: The STD consequences of adolescent virginity pledges. *Journal of Adolescent Health, 36,* 271–278.

Chatters, L. M., Taylor, R. J., & Jayakody, R. (1994). Fictive kinship relations in Black extended families. *Journal of Comparative Family Studies, 25,* 297–312.

Cherlin, A. J., & Furstenberg, F. F. (1994). Stepfamilies in the US: A reconsideration. *Annual Review of Sociology, 20,* 359–381.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology, 94,* S95–S120.

Crosnoe, R. (2004). Social capital and the interplay of families and schools. *Journal of Marriage and Family, 66,* 267–280.

Darwin, C. (1859). *On the origin of species*. http://www.literature.org/authors/darwin-charles/the-origin-of-species/.

Furstenberg, F. F. (1987). Then new extended family: The experience of parents and children after remarriage. In K. Pasley & M. Inhinger-Tallman (Eds.), *Remarriage and stepparenting: Current research and theory* (pp. 42–61). New York: Guilford Press.

Furstenberg, F. F. (2005). Banking on families: How families generate and distribute social capital. *Journal of Marriage and Family, 67,* 809–821.

Goodall, J. (1988). *In the shadow of man*. New York: Houghton Mifflin.

Gould, S. J. (1988). *Introduction to the revised edition. In Jane Goodall, In the shadow of man* (pp. v–viii). New York: Houghton Mifflin.

Guo, G., Roettger, M. E., & Cai, T. (2008). The integration of genetic propensities into social-control models of delinquency and violence among male youths. *American Sociological Review, 73,* 543–568.

Hinde, R. A. (1987). *Individuals, relationships and culture*. Cambridge: Cambridge University Press.

Just, M. A., & Varma, S. (2007). The organization of thinking: What functional brain imaging reveals about the neuroarchitecture of complex cognition. *Cognitive, Affective, & Behavioral Neuroscience, 7,* 153–191.

Knoke, D., & Bohrnstadt, G. W. (1991). *Basic social statistics*. Itasca, Il: F. E. Peacock.

Landau, E. (2012). Rover finds evidence of ancient water on Mars. http://www.cnn.com/2012/09/27/us/mars-rover-water/index.html.

Lerner, P. N. (1996). *Handbook of ethological methods*. Cambridge: Cambridge University Press.
Levin, I. (1993). Family as mapped realities. *Journal of Family Issues, 14,* 82–91.
Lucia, V. C., & Breslau, N. (2006). Family cohesion and children's behavior problems: A longitudinal investigation. *Psychiatry Research, 141,* 141–149.
Miller, G. E., Lachman, M. E., Chen, E., Gruenewald, T. L., Karlamangla, A. S., & Seeman, T. E. (2011). Pathways to resilience: Maternal nurturance as a buffer against the effects of childhood poverty on metabolic syndrome at midlife. *Psychological Science, 22,* 1591–1599.
Moos, R., & Moos, B. (1994). *Family environment scale manual: Development, applications, research* (3rd ed.). Palo Alto: Consulting Psychologist Press.
Olson, D. H. (2000). The Circumplex Model of marital and family systems. *Journal of Family Therapy, 22,* 144–167.
Olson, D. (2011). FACES IV and the Circumplex Model: Validation study. *Journal of Marital & Family Therapy, 3,* 64–80.
Putnam, R. (2000). *Bowling alone: The collapse and revival of American community*. New York: Simon and Schuster.
Ravanera, Z. R., & Rajulton, F. (2010). Measuring social capital and its differentials by family structures. *Social Indicator Research, 95,* 63–89.
Sandefur, G. D., Meier, A. M., & Campbell, M. E. (2006). Family resources, social capital, and college attendance. *Social Science Research, 35,* 525–553.
Sarkisian, N., & Gerstel, N. (2012). *Nuclear family values, extended family lives: The importance of gender, race, and class*. New York: Routledge.
Scott, W. A., & Scott, R. (1981). Intercorrelations among structural properties of primary groups. *Journal of Personality and Social Psychology, 41,* 279–292.
Simons, R. L., Lei, M. K., Beach, S. R. H., Brody, G. H., Philibert, R. A., & Gibbons, X., F (2011). Social environment, genes, and aggression: Evidence supporting the differential susceptibility perspective. *American Sociological Review, 76,* 883–912.
Slatcher, R. B., Robles, T. F., Repetti, R. L., & Fellows, M. D. (2010). Momentary work worries, marital disclosure, and salivary cortisol among parents of young children. *Psychosomatic Medicine, 72,* 887–896.
Smetana, J. G. (1988). Concepts of self and social convention: Adolescents' and parents' reasoning about hypothetical and actual family conflicts. In M. R. Gunnar, & W. A. Collins (Eds.), *Development during the transition to adolescence. Minnesota Symposia on Child Psychology* (Vol. 21, pp. 79–122). Hillsdale: Lawrence Erlbaum Associates.
U. S. Census Bureau(2012). (http://www.census.gov/cps/about/cpsdef.html)
Widmer, E. D. (1999). Family contexts as cognitive networks: A structural approach of family relationships. *Personal Relationships, 6,* 487–503.
Widmer, E. D. (2006). Who are my family members? Bridging and binding social capital in family configurations. *Journal of Social and Personal Relationships, 23,* 979–998.
Wilson, E. O. (2012). *The social conquest of the Earth*. New York: Liveright.
Wise, R. A., & King, A. R. (2007). Family environment as a predictor of the quality of college students' friendships. *Journal of Family Issues, 29,* 828–848.

Part IV
Family Programs and Policies

Chapter 12
Multiple Levels and Modalities of Measurement in a Population-Based Approach to Improving Parenting

Carol W. Metzler, Matthew R. Sanders and Julie C. Rusby

The need for a public health approach to strengthening parenting through wide-reaching evidence-based parenting supports is gaining recognition worldwide (Sanders 2010). The goal of such an approach is to increase the prevalence of effective parenting in the entire population. Accomplishment of this goal requires that a large proportion of the population is reached with effective parenting strategies (Biglan 1995), which need to be delivered in a variety of formats, across a variety of settings, and at different levels of intensity so that they are widely accessible in the community. Targeting a goal of changing the population prevalence of an outcome forces us to consider issues of a program's reach, efficacy, adoption rate, quality of implementation, and length of maintenance (Glasgow et al. 1999); adopting this orientation will move the research on family interventions forward, and will move practice and policy forward as well.

To facilitate accountability and inform program evaluation, practice, and policy, a public health approach requires a spectrum of measurement tools, so that outcomes of an intervention can be assessed at multiple levels. These levels include (1) micro-focused measures that obtain precise, objective measures of parenting and child behavior; (2) valid survey measures of parenting and child behavior that can be inexpensively administered; (3) population-level indicators of the prevalence of specific outcomes; and (4) measures of the implementation of a program in an applied setting. This chapter will discuss each of these levels of measurement and the contribution that each makes as a building block toward a public health strategy, in which the goal is to find practical, affordable, timely, and efficient ways to collect and meaningfully summarize valid data for real-world utility.

C. W. Metzler (✉) · J. C. Rusby
Oregon Research Institute, Eugene, OR, USA
e-mail: carolm@ori.org

M. R. Sanders
Parenting and Family Support Centre, University of Queensland, Brisbane, QLD, Australia
e-mail: matts@psy.uq.edu.au

J. C. Rusby
e-mail: juliecr@ori.org

S. M. McHale et al. (eds.), *Emerging Methods in Family Research,*
National Symposium on Family Issues 4, DOI 10.1007/978-3-319-01562-0_12,

Specifically, observation measures are essential for objectively establishing the evidence base of parenting interventions; they also have significant utility as a measure of family change in a clinical setting. This chapter describes micro-social direct observation methods for observing parenting and child behaviors, how play tasks might be designed to elicit specific kinds of parent and child behaviors, and the utility of this approach in a public health strategy. Second, self-report surveys of parents are explored, with an emphasis on the survey features critical for success in a public health approach. Three parent survey measures that meet these criteria are highlighted. The third section of the chapter focuses on population-level indicators for assessing changes in the prevalence of parenting and child outcomes, such as child maltreatment rates, special education referrals, and surveillance surveys. The benefits as well as challenges inherent in using these measures are discussed. Finally, the chapter discusses implementation measures as an essential component of evaluating a population rollout. Discussion focuses on how each of these levels of measurement contributes to a public health strategy for improving parenting and child outcomes, whether in the context of a randomized controlled trial or a population rollout of a program.

Micro-Social Measures of Child Behavior and Parenting

The first level of data to be highlighted is microsocial measures of parents' and children's behavior in interaction with each other, in which parents' and children's interaction behaviors and affect are observed and coded by independent observers. Research based on coded exchanges between parents and children have contributed to theories, intervention development, and clinical practice for a wide variety of child behavior problems, including coercive processes in families of children with oppositional and aggressive behaviors (Patterson et al. 1992), difficult interactions during structured tasks for children with ADHD (Danforth et al. 1991), the importance of parents' cognitive scaffolding to facilitate children's learning (Fagot and Gauvain 1997), and the impact of parents' proactive strategies for children with conduct problems (Gardner et al. 1999).

Direct observation of behaviors offers greater objectivity than self-reports, minimizing subjective biases due to mood (Fergusson et al. 1993; Müller et al. 2011) or memory (Gorin and Stone 2001; Schwarz and Sudman 1994). Structured observation assessments of mother-child interactions have been found to better predict children's social skills and academic achievement (Zaslow et al. 2006) and better detect parental maltreatment (Bennett et al. 2006) than mothers' ratings of their own parenting. Microsocial coding of sequences of parent-child interactions can identify specific antecedents that "trigger" problematic behaviors, consequences that reinforce them, and the reciprocal influences between parents and children. Microsocial measures tend to be more sensitive to change than molar-level questionnaire report data, and thus are advantageous for assessing change in randomized controlled trials and in clinical settings (e.g., Bor et al. 2002; Gardner et al. 2006; Webster-Stratton and Hammond 1997; Wells et al. 2006). Such sensitivity begets challenges, however, such as large amounts of variance in the data (Stoolmiller et al. 2000) due to differences in settings, activities, and people present.

The Observation Setting and Structure of Activities

Selecting observation procedures that match the outcomes of interest is critical for validity and utility of the results. Although non-structured observations in naturalistic environments, such as the home, are the most ecologically valid, parameters frequently need to be established (such as restricting television watching and visitors) to ensure that the target persons interact during the observation period (e.g., Gardner et al. 2003). An alternative method is to conduct a lab-task (or play-task) assessment that involves a standardized set of toys or materials with standard prompts for activities in which a parent and child engage.

For both clinical and public health approaches, semi-structured play-task observations provide a number of advantages over home observations. First, play tasks are more practical as they can be conducted in a lab or clinic setting. Second, play tasks can be designed to emulate real-life situations and elicit specific behaviors so that low base-rate behaviors can be assessed more efficiently. Third, the standardized procedures and materials used in play tasks create consistency across participants and assessment timepoints, thus decreasing variability found in naturalistic settings. Such standardization allows for comparisons between participants (e.g., for comparing subgroups) and within participants (for monitoring progress toward intervention goals). Play tasks are not entirely free of error variance due to children's differential familiarity with the toys, reactions to new places, and difficulty with the activities. To resolve this, a series of different tasks are often conducted in order to account for task-specific variance.

Studies comparing behavior observed in unstructured home settings vs. semi-structured play-task lab settings have found that behaviors observed in the lab correspond to behavior observed in the home for both parents and children (Dadds and Sanders 1992; Webster-Stratton 1985), but that play tasks tend to elicit more parent directives and praise and fewer child conduct problems than unstructured activities in the home (Webster-Stratton 1985). Thus, generalizing behaviors observed in the lab to the home should be done with caution (Gardner 2000).

Examples of Play Tasks and Target Behaviors

Typically, specific play tasks are selected to emulate real-life situations and to elicit target behaviors. For example, clean-up tasks, in which the parent interrupts the child's play to tell them to put the toys away, are often used to measure parent-child conflict, parent directives, and child compliance/non-compliance to parent directives (e.g., Gardner et al. 1999; Dishion et al. 2008). Teaching tasks involve the parent teaching or guiding the child to do a novel task, eliciting the ways in which parents provide cognitive assistance to their children (Fagot and Gauvain 1997).

Tasks in which the parent is kept busy (e.g., filling out a form or questionnaire) and unable to attend fully to the child have been used to assess parents' strategies when preoccupied with a task while their child's time is unstructured (e.g., Gardner et al. 2006). Problem-solving tasks involve parent(s) and child attempting to solve problems that they identify having, such as doing chores or completing homework (Dadds and Sanders 1992; Degarmo and Forgatch 2012). Planning a fun activity together has been used to assess family communication styles and affect (e.g., Scott et al. 2011; Stoolmiller and Snyder 2004). These examples illustrate that in order to select ecologically valid play-task scenarios, it is essential to identify the parent and child behaviors of interest and the "real-life" situations in which those behaviors tend to occur, then to select play-task scenarios that can emulate those real situations.

Example: The Parent-Child Play Task and Observation System (PCPTOS).

Although it is generally assumed that structured play tasks are effective at eliciting the behaviors of interest, to our knowledge this assumption has not been directly tested with a sufficiently sized sample. Using baseline data from a randomized controlled trial of a video-based parenting program (Metzler et al. 2012), the authors examined the extent to which a series of structured play tasks for mother-child dyads elicited the parent and child behaviors of interest. Mothers ($N = 322$) and their 3- to 6-year-old children with oppositional and/or aggressive behavior problems participated. The play-task assessment involved: (1) a 20-min Mother-Busy task in which the mother completed a questionnaire while the child was given a picture book, to examine child behavior when mother is otherwise occupied as well as mother response to child's interruptions and demands; (2) a 10-min Free-Play task with a standardized toy set, to examine spontaneous mother-child interactions; (3) a Clean-Up task (up to 15 min), to elicit mother directives and examine child response to directives; (4) a 10-min Teaching task with a puzzle above the child's age level, to elicit mother's use of cognitive assistance (guidance); and (5) a 10-min Transition task (from Free-Play to Clean-Up to Mother-Busy), to examine mother preparation for transitions, child reactions to transitions, and end-of-session fatigue.

The play tasks were video-taped and coded by independent coders. For each task, the frequency of specific behaviors was compared for the task designed to elicit that behavior vs. all other tasks. For example, the frequency of child interruptions and demands during the Mother-Busy task was compared to the frequency of these behaviors during the other four tasks. The first four tasks were found to successfully elicit the behaviors of interest: Greater rates of child interruptions ($t = 26.95$, $p < 0.001$) and demands ($t = 6.88$, $p < 0.001$) occurred during the Mother-Busy task; greater rates of neutral mother conversing ($t = 15.98$, $p < 0.001$) and child conversing ($t = 3.11$, $p = 0.002$) occurred during Free-Play; and more guidance was provided by mothers during the Teaching task ($t = 3.71$, $p < 0.001$). The Clean-Up task elicited more positive ($t = 17.34$, $p < 0.001$), neutral ($t = 10.17$, $p < 0.001$), and negative

($t = 7.97$, $p < 0.001$) directives from mothers, as well as more child compliance ($t = 17.95$, $p < 0.001$) and non-compliance ($t = 9.66$, $p < 0.001$). The Transition task did not elicit greater child negative affect as was expected, indicating that children were not distressed from fatigue at the end of the session. These results demonstrate how structured play tasks can elicit specific target behaviors, allowing researchers, evaluators, and clinicians to objectively assess parent and child behaviors that may otherwise occur infrequently and/or be difficult to capture in natural situations.

Important Considerations Although this example involves mother-child dyads, the research or clinical goals and the theoretical approach underlying the assessment questions will guide the decision of who to include in the play-task assessments–mother, father, both parents, siblings (Kerig 2001). Before launching a research study or clinical procedures involving a structured play-task assessment, it is important to pilot test the selected play task procedures to see how well the activities, materials, and duration work with the target participants, as well as pilot test the selected method for summarizing the interactions (e.g., the microsocial coding system).

Utility Within a Public Health Framework

Microsocial assessment systems can be used to measure outcomes in randomized controlled trials, to monitor families' progress in clinical settings, or if collected on a representative subsample of a broader population, to summarize changes at the population level. The primary advantages of microsocial assessment systems are the objectivity of the data, the ability to conduct sequential analyses of interaction patterns (Bakeman and Gottman 1997; Patterson et al. 1992), and greater sensitivity than questionnaire measures (Snyder and Stoolmiller 2002).

Despite these advantages, microsocial assessment systems present significant challenges for clinical or public health approaches in finding practical, affordable, timely, and efficient ways to collect and meaningfully summarize the data for real-world utility. To collect and code observational data reliably can require lengthy training for coders and a computerized system or coding device. These coding systems result in complex data that then need to be analyzed or summarized into meaningful reports, and the methodologies for analyzing observational data are themselves quite complex (e.g., Granic and Lamay 2002; Stoolmiller et al. 2000; Stoolmiller and Snyder 2004). The development of easy-to-use computerized data report systems that can provide immediate feedback in an easy-to-interpret graphic format are an important next step for enabling broader utilization of microsocial measures to evaluate population-level parenting interventions. There is a great need to harness technology to reduce costs and increase the reach of these measures. Currently, microsocial measures are used primarily in research studies, or as a gold standard from which to develop and validate self-report surveys, which are often more feasible to collect in community settings (Hawes and Dadds 2006; Scott et al. 2011).

Self-Report Measures of Child Behavior and Parenting

The second level of data to be highlighted is parents' self-report measures of child behavior and parenting practices. In large-scale implementations of parenting interventions, self-report measures are required for routine use in evaluating individual- or family-level client outcomes. To serve this purpose, these measures must be reliable, valid, and sensitive to changes in the child behaviors and parenting practices targeted by a given intervention. Most well-established evidence-based parenting programs recommend that a set of parent-completed measures be routinely used to assess clinical outcomes. For example, a useful clinical assessment package could include measures of sociodemographics of the participating parent (e.g., *Family Background Questionnaire*, Sanders et al. 1996), child behavior and adjustment (e.g., *Strengths and Difficulties Questionnaire*, Goodman 1997; *Eyberg Child Behavior Inventory*, Eyberg and Pincus 1999), parenting practices (e.g., *Parenting Scale,* Arnold et al. 1993; *Brief Alabama Parenting Questionnaire*, Shelton et al. 1996), parental emotional distress (*Depression, Anxiety, and Stress Scale*, Lovibond and Lovibond 1995), family conflict (*Parent Problem Checklist*, Dadds and Powell 1991), and a consumer satisfaction measure (*Consumer Satisfaction Questionnaire*, Sanders et al. 1996). This assessment package is described in more detail in various practitioner manuals for the Triple P Positive Parenting Program (e.g., Sanders et al. 1998). Unfortunately, the routine use of such standardized instruments to assess family-level outcomes after delivery of a parenting program is not currently the norm at most clinical service agencies. Thus, one cannot assume that service agencies have well-developed monitoring and tracking systems for assessing outcomes. Program developers, implementers, and evaluation teams need to encourage and assist service agencies in adopting outcome assessment tools and embedding them in routine client evaluations. Parenting practitioners often need training in how to administer and interpret the assessment packages, how to discuss them with clients, and how data collection efforts contribute to an overall evaluation of a program when data from multiple agencies are combined.

Desirable Features of Self-Report Measures in a Public Health Approach to Parenting

To be maximally useful in a population-level approach to improving parenting, self-report outcome measures should have a number of key features. First, they need to be easy to administer in a variety of formats (e.g., paper-pencil, online, interview). Second, they are more likely to be accepted by agencies if they are in the public domain than if distributed only through test publishers, due to cost considerations. Third, measures should have consistent, simple scaling and response options to ensure greater consistency and to reduce parental confusion associated with having a myriad of different response formats within the same questionnaire.

Fourth, items should have good face validity for ease of interpretation, have good internal consistency and test-retest reliability, and be comprehensive enough to provide valid indices of targeted child outcomes (e.g. conduct and internalizing problems, prosocial behavior) as well as family risk and protective factors related to these child problems (positive and coercive parenting practices, parental stress and depression, family conflict). The purpose of each measure needs to be readily apparent and acceptable as well: Practitioners need to be able to see the usefulness of the information gathered from each measure and be able to provide brief, plausible explanations to parents about its usefulness as well, in order for parents to be willing to complete it.

In addition, because parenting interventions are used across a range of settings with families who have varying levels of risk, measures need to be sensitive to changes in both clinical and non-clinical populations. Similarly, a measurement system is more beneficial if it can detect clinically meaningful change across multiple levels of an intervention, from brief/ low-intensity to more intensive. For example, in a number of large-scale population roll-outs, the same set of outcome measures has been used to assess the effects of brief, low-intensity primary care interventions to more intensive clinical interventions (e.g., Sanders et al. 2008). Finally, measures that have been translated into multiple languages and shown to be acceptable and retain sound psychometric properties across different languages and cultures are particularly useful.

There are surprisingly few measures that meet all of these criteria. Many self-report measures are too long to be practical; others are proprietary and thus too costly for wide-scale use. Many focus too narrowly on one or two domains and are not comprehensive enough in the range of outcomes that they measure; others are too complex for straightforward interpretation and analysis. Finally, few are available in different languages. Three notable examples of measures that do generally meet these criteria are described below.

Examples of Measures Used in Evaluations of Large-Scale Parenting Interventions

The *Strengths and Difficulties Questionnaire (SDQ,* Goodman 1999) can be used to identify behavioral and emotional problems in children and adolescents. It has been used in many countries in clinical and epidemiological contexts, as well as in the evaluation of several population-based parenting interventions (e.g. Sanders et al. 2008; Nexus Management Consulting 2011). The *SDQ* has five subscales of five items each: conduct problems, hyperactivity, emotional problems, peer problems, and prosocial behavior. Parent and teacher versions are provided for children ages 3–16, and a self-report version for youth ages 11–16. Extended versions of the parent and teacher *SDQ* include an impact supplement that asks if the child has a problem with emotions, concentration, behavior, or getting along with people, and if yes, then the degree of chronicity, distress, social impairment, and burden to others (Goodman 1999). The *SDQ* is brief (approximately 10 min), widely accessible (www.sdqinfo.com), and has norms from many countries and languages. It is available in English, Spanish, French, German, Dutch, Portuguese, Italian, Swedish, Finnish, Croatian, Arabic, and Urdu. In addition, there are no fees or permissions required for using the paper-and-pencil version, nor does it require special training and/or professional or academic credentials to administer.

Scores discriminate well between low- and high-risk samples (Goodman and Scott 1999). The Total Difficulties score has adequate internal reliability ($\alpha = 0.76$) and test retest reliability ($r = 0.85$). The *SDQ* correlates well with another widely used but lengthier instrument, the *Child Behavior Checklist* (Achenbach 1991), and is often preferred by parents (Goodman and Scott 1999). It is important to note, however, that the *SDQ* was developed as an epidemiology screening tool and not as an instrument to test intervention effects. For example, the time interval that parents and teachers are asked to consider when rating the child's behavior (6 months) is appropriate for epidemiological surveys but too long for parenting intervention studies, in which both short- and long-term effects need to be assessed. In intervention trials, the follow-up version of the *SDQ* is typically used, in which the six-month time period is reduced to 1 month, to avoid encompassing the pre-intervention period as well as the post-intervention period.

The *Child Adjustment and Parent Efficacy Scale* (*CAPES*, Morawska and Sanders 2010) is a new measure of child behavioral and emotional adjustment and parental efficacy. It consists of 30 items rated on a 4-point scale, from "not true of my child at all" (0) to "true of my child very much, or most of the time" (3); higher scores indicate greater problems. Twenty-six items assess child behavior concerns (e.g., My child rudely answers back to me) and behavioral competencies (Behavior subscale; e.g., My child accepts rules and limits); and four items assess emotional adjustment (Emotional Adjustment subscale; e.g., My child worries). A Total scale score can also be computed. The Behavior, Emotional Adjustment, and Total scales all show high internal consistency ($\alpha = 0.91$, 0.81, and 0.91, respectively). The Efficacy subscale asks parents to rate their level of confidence in managing 20 child emotional and behavioral problems. Items are rated on a 10-point scale, from "certain I can't do it" (1) to "certain I can do it" (10), with higher scores indicating greater confidence. The internal consistency for these items is high ($\alpha = 0.96$).

The *Parenting and Family Adjustment Scale* (*PAFAS*, Sanders and Morawska 2010) is a 40-item measure designed to assess aspects of family functioning. Parents rate items on a four-point scale (0 = "not true of me at all," to 3 = "true of me very much, or most of the time"). The *PAFAS* has four subscales, each yielding a separate score: Parenting Practices (28 items), Parent Adjustment (5 items), Family Relationships (4 items), and Parental Teamwork (3 items). All subscales show good internal consistency ($\alpha = 0.76$, 0.86, 0.78, and 0.74 respectively). In addition, a Total score involving all 40 items can be derived that shows good internal consistency ($\alpha = 0.88$).

The *CAPES* and the *PAFAS* were both specifically developed to create parenting/family intervention evaluation tools that have many of the desirable features described above. The *CAPES* and *PAFAS* were designed to cover key child and family domains in two relatively brief, easy-to-administer instruments. They are in the public domain, and the items are face valid and psychometrically sound, with consistent response formats. They are both currently being translated into multiple languages and being validated in studies across multiple countries. Future use in evaluation studies will provide needed information about their sensitivity to change in various populations.

Important Considerations

As with microsocial measures, the decision of which family members should be asked to provide data is an important consideration. For example, who should report on the child's behavior and family-level indicators? If only one parent is selected, should it be the "primary parent"? How is primary parent defined, especially in non-traditional family structures? Should both mothers and fathers be asked to provide reports, and if so, how will discrepancies in their reports be resolved, especially if one parent spends substantially more time in the parenting role? Should step-parents and non-married co-parenting partners provide data as well, and how should their data be handled? These questions must be answered by those conducting the evaluation, guided by the focus of the intervention and the goals of the evaluation. Generally speaking, agreement among reporters on child behaviors and family-level indicators, even those in the same household, should not necessarily be expected; it is often most useful to view each reporter as having a unique perspective to contribute to a more complete understanding of the family.

Utility Within a Public Health Framework

Like micro-social assessments, parent-report questionnaires can be used for a variety of purposes–to measure outcomes in randomized trials, to monitor individual families' clinical progress, and to assess changes across broad populations. Unlike microsocial measures, however, brief parent-report surveys in the public domain can be administered and analyzed relatively inexpensively across a broad population of families in a community or region, providing much more utility in a public health rollout of a parenting intervention than more expensive measures. The primary disadvantage of parent-report measures is that they are not objective and can be vulnerable to parents' social desirability, self-perception, and memory biases. They are also more global and less detailed than microsocial measures. Despite this, the relative ease of administration and low cost of psychometrically sound, public-domain parent-report measures make them important research, clinical, and public health assessment tools.

Population-Level Indicators of Child and Parenting Outcomes

The third level of data to be highlighted is population-level indicators, such as archival administrative records and population surveillance surveys. Large-scale parenting interventions that aim to change the prevalence of target outcomes require measures that focus at the population level. In addition, monitoring key child, adolescent, and family indicators at the societal level can provide a community, county, or state with accurate estimates of the well-being of the entire population of young people, and thus can guide decision making regarding programs, policies, and practices to improve child/youth well-being (Mrazek et al. 2003). Publicly available archival measures that can be used to evaluate prevalence rate changes in child and youth well-being include records of substantiated child maltreatment, child out-of-home placements, child hospitalizations and emergency room visits, child deaths, infant mortality, mental health services utilization, suicide attempts, high school dropout, grade-level proficiency scores, school suspensions/expulsions, special education referrals, teen pregnancy, juvenile arrests, juvenile incarceration, and crimes against persons and/or property. In addition, large-scale household surveys can be used to determine the prevalence of child behavioral/emotional difficulties and family risk factors (Sanders et al. 2007). Large-scale student surveys in schools can be used to determine the prevalence of healthy and harmful behaviors in youth as well as family risk and protective factors (Metzler et al. 2008; Mrazek et al. 2003). Unfortunately, few studies to date have employed population-level measures to assess the effects of large-scale parenting interventions.

Examples of the Use of Population-Level Indicators

Child Maltreatment Three population indicators related to child maltreatment were used by Prinz et al. (2009) to evaluate the effects of a population-level parenting intervention trial in South Carolina. These indicators were derived from independent data collection systems, deposited into a state-run statistical division. First, substantiated child maltreatment (CM) data, recorded by Child Protective Services (CPS) staff, were utilized. These data were unduplicated, such that no CM case was counted more than once in a given year. The second indicator was child out-of-home placements, recorded through the foster care system. Third, data on child hospitalizations and emergency room visits due to CM injuries were utilized, recorded by medical staff in compliance with mandatory state reporting requirements for hospitals. All three population-level indicators were computed as annual rates per 1,000 children in the birth to 8-year-old age range.

Children's Mental Health Computer-Assisted Telephone Interview techniques were employed by Sanders et al. (2007) and Sanders et al. (2008) to establish the base prevalence of children's behavioral and emotional problems and of target parenting/family risk and protective factors in the state of Queensland, Australia (population

4.6 million). The survey design involved deriving a random sample of private households and interviewing the primary caregiver in households with at least one child 12 or younger. A simple random sample of listed private telephone numbers was drawn from a database of all listed private numbers in Queensland. This sample was supplemented with a random sample of unlisted numbers in the Brisbane and Gold Coast area of Queensland, where the proportion of unlisted numbers is higher than in the rest of the state (more than 15 %). This scheme produced a sample of numbers that approximated a simple random sample of households with a fixed telephone. A small proportion of the target population (approximately 4 % of all Queensland households) was excluded because they did not have a fixed telephone at the time.

Readiness and Success in School The *Australian Early Development Index* (*AEDI*), assessed every 3 years in a national sample of approximately 270,000 children in their first year of full-time school, is an example of a population measure of young children's development at a key point of developmental transition. Like a census, it involves creating a snapshot of children's development in communities across Australia. Teachers complete a checklist for children in their first year of full-time school; the checklist measures five key domains of early childhood development: physical health, social competence, emotional maturity, language and cognitive skills, and communication skills and general knowledge. Data from the *AEDI* has been used to support community initiatives to foster healthy child development; facilitate more effective use of local resources; strengthen links between schools, preschools, local government, healthcare, and other local organizations; assist schools with planning for optimal transition for incoming students; and re-focus community services and systems towards children.

Child and Youth Well-Being The *Youth Risk Behavior Surveillance System* (*YRBSS*; www.cdc.gov/yrbs/) is a large-scale youth survey funded by the Centers for Disease Control and Prevention (CDC) to track the prevalence of various healthy and harmful behaviors in youth, including substance use, unsafe sexual behaviors, behaviors that contribute to unintentional injuries and violence, diet, and physical activity. The *YRBSS* is conducted through school-based surveys of representative samples of 9th–12th grade students. It involves a national-level survey conducted by the CDC, as well as state-, tribal-, and local-level surveys conducted by state and local education and health agencies and tribal governments. These data can be used to assess trends at the national as well as the local level, evaluate the impact of population-level program and policy initiatives, and assess progress toward community, county, and state goals.

Issues in Using Population-Level Indicators

Key issues remain in using population-level indicators to evaluate a public health approach to parenting intervention. First, one concern with employing archival data alone to establish changes in the prevalence of an outcome, such as child maltreatment, is that the criteria and reporting processes for identifying a "child maltreatment case" can vary considerably across administrative systems, creating inconsistent estimates of cases across regions. Thus, developing more reliable and widely agreed-upon criteria to be used by child welfare officials for recording instances of child maltreatment would be beneficial. Also, the utility of archival data is stronger when independently collected sources of data can be combined (e.g., using rates of founded child maltreatment cases, of hospitalization due to maltreatment injuries, and of out-of-home placements to measure the prevalence of child maltreatment).

Second, population-level parenting interventions would benefit from data linkage, a process of connecting information across two or more different sources of data that relate to the same person, family, place, or event. Administrative information is created each time a person comes into contact with a particular agency, such as a birth or death registration, hospital stay, emergency room visit, child protective services visit, or arrest, but these records are not typically associated with the same identification number across agencies. Connecting these pieces of information for the whole population in ways that do not breach an individual's privacy would enable a much richer analysis when evaluating a community's health. The value of linking birth records to child protection data in order to track population trends in the presence of risk factors for child maltreatment was shown by Putnam-Hornstein et al. (2011). The advantages of data linkage for evaluating prevention programs include: (a) adding value to and generating a research return on the substantial existing investment in routine administrative and clinical data sets; (b) integrating data from one sector (e.g. health) with data from a wide range of other sectors (e.g., education, law enforcement), without requiring unique personal identifiers across sectors; (c) protecting families' privacy by reducing the need for release of names and other personal identifiers to researchers; (d) fostering research among parenting, public health, clinical, and biomedical researchers, with direct benefits for clinical outcomes; and (e) promoting community development efforts through enhanced interactions among researchers, clinicians, administrators, consumer groups, and the mass media.

The data linkages required to evaluate a population-level parenting intervention would vary depending on the focus of the intervention and the outcomes a program sought to achieve. For example, an intervention aimed at improving children's socioemotional skills may wish to connect education outcomes (e.g., academic attainments, suspensions, special education referrals) with child maltreatment data. An intervention focused on older children with conduct problems may wish to link data from juvenile justice, police, education, and mental health services.

Measures of the Implementation of Parenting Programs

The fourth level of data to be highlighted is measures of the implementation of parenting programs. Implementation research underscores the importance of effective implementation of evidence-based practices, because desirable outcomes are achieved only when effective programs are implemented well (Fixsen et al. 2005).

Measures to track the adoption, implementation, maintenance, and reach of a parenting intervention are essential to evaluating the degree to which the public rollout of a program is successful and the extent of its public health impact (Glasgow et al. 1999). In addition, the program start-up process is recursive, with many iterative stages and steps that are not well understood (e.g., engagement, readiness planning, staff hiring and training). Valid implementation measures can begin to better elucidate this process (Saldana et al. 2011). Constructs to be measured include acceptability, appropriateness, and feasibility of the program to stakeholders, consumers, and particular settings; adoption of the program within the agency; fidelity of its implementation; its penetration or reach; sustainability or maintenance; and its cost of implementation (Fixsen et al. 2005; Proctor et al. 2011).

Examples of Measures of Implementation

The measurement of implementation is still in its infancy, and few validated measures exist. The *Stages of Implementation Completion* (*SIC*; Saldana et al. 2011) is a new measure of the adoption and implementation process that breaks the process down into eight stages (e.g., engagement, consideration of feasibility, readiness planning, staff hired and trained), with subactivities within each stage (e.g., within "staff hired and trained": date agency checklist completed, date first staff hired, date clinical training held). The completion date for each activity is entered, providing scores for rate of completion, proportion of activities completed, and duration of each stage. In a dissemination trial for Multidimensional Treatment Foster Care (Chamberlain 2003), the proportion of activities completed and the duration of time in stages 1–3 on the *SIC* both predicted successful start-up of the program at stage 6 (Saldana et al. 2011).

Spoth and colleagues used observations of intervention adherence to measure quality of implementation in the PROSPER study (Spoth et al. 2011), a dissemination trial of family-based (Strengthening Families 10–14) and school-based (Life Skills Training, Project ALERT) interventions. In this measure, live observers attended intervention sessions and completed adherence checklists, yielding a score for proportion of program-specific content actually delivered. Observers also completed Likert four-point ratings of participants' engagement and participation, as well as facilitation quality (friendliness, answering questions effectively) for the family intervention. Results showed that high levels of implementation quality were maintained across the 6 years of the study, which enabled examination of the correlates of effective implementation (Spoth et al. 2011).

Utility of Measuring Implementation

Implementation research is a field still in its infancy with no "home" discipline, resulting in studies scattered across disciplines and journals, with inconsistent use of language and constructs. Despite this, or perhaps because of it, the development of valid measures of implementation process and quality are critical for advancing the field of implementation science. Few implementation measures currently exist, and fewer still have been validated. The development of such measures, however, will build our understanding of the implementation process and allow us to identify behaviors and activities that lead to successful implementation. Valid measures of implementation will also allow us to model implementation success as a function of combinations of program factors (acceptability, effectiveness), agency or setting factors (feasibility), and implementation factors (adherence, cost, sustainability), leading to a much richer understanding of how to foster implementation success, guide implementation decisions (Proctor et al. 2011), and measure public health impact (Glasgow et al. 1999).

Summary and Conclusions

In this chapter, we have reviewed four levels and modalities of measurement for evaluating the impact of a parenting program: microsocial observational measures, self-report survey measures, population-level archival and surveillance measures, and implementation measures. Each of these levels of measurement contributes to building a public health strategy for improving parenting and child outcomes, and each comes with its own advantages and challenges with regard to sensitivity, specificity, objectivity, feasibility, cost, and scale. Implementers and evaluators must choose the best balance of measures to match the goals, needs, resources, and scope of their planned evaluation. A focus on the clinical outcomes of an agency's adoption of a parenting intervention may include microsocial, self-report, and implementation measures but exclude population-level measures, while a community-wide population-level roll-out of a program may include self-report, population-level, and implementation measures but exclude microsocial measures. A randomized controlled trial of a population roll-out, however, may demand additional rigor and bring to bear additional resources that would allow a combination of all four levels of measurement.

The field of family interventions has made great strides in developing valid observational and self-report measures of parenting and child behavior, and in identifying archival and surveillance measures that provide population-level data on child and family well-being. Many challenges remain, however, including the high cost of microsocial measures, the need for better use of technology to improve their reach and reduce their costs, low concurrence between observed behaviors and self-reports, difficulties linking data across methods and sources, and challenges in minimizing sampling bias with regard to fathers, minorities, and one vs. two parents. In addition, much work remains to develop valid measures of implementation, as this field is very young. Addressing all of these challenges will further build our ability to improve child and family well-being.

Within a public health framework, the most successful measures will be valid but also relatively straightforward to analyze and interpret, especially if not part of a scientific study. Successful measures are understandable and interpretable to clinicians, government officials, the media, community leaders, and the general public, in order to guide decision-making, provide accountability, and generate support for parenting interventions.

Acknowledgments Preparation of this chapter was funded by grant #R01DA021307 from the National Institute on Drug Abuse to Carol W. Metzler at Oregon Research Institute.

References

Achenbach, T. M. (1991). *Manual for the child behaviour checklist/4–18 and 1991 profile*. Burlington, VT: University of Vermont Department of Psychiatry.

Arnold, D. S., O'Leary, S. G., Wolff, L. S., & Acher, M. M. (1993). The parenting scale: A measure of dysfunctional parenting in discipline situations. *Psychological Assessment, 5,* 137–144.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis (2nd ed.)*. New York, NY: Cambridge University Press.

Bennett, D. S., Sullivan, M. W., & Lewis, M. (2006). Relations of parental report and observation of parenting to maltreatment history. *Child Maltreatment, 11,* 63–75.

Biglan, A. (1995). Translating what we know about the context for anti-social behavior into a lower prevalence of such behavior. *Journal of Applied Behavior Analysis, 28,* 479–492.

Bor, W., Sanders, M. R., & Markie-Dadds, C. (2002). The effects of the Triple P-Positive parenting program on preschool children with co-occurring disruptive behavior and attentional/hyperactive difficulties. *Journal of Abnormal Child Psychology, 30,* 571–587.

Chamberlain, P. (2003). The oregon multidimensional treatment foster care model: Features, outcomes, and progress in dissemination. In S. Schoenwald & S. Henggeler (Eds.). Moving evidence-based treatments from the laboratory into clinical practice. *Cognitive and Behavioral Practice*, *10,* 303–312.

Dadds, M. R., & Powell, M. B. (1991). The relationship of interparental conflict and global marital adjustment to aggression, anxiety, and immaturity in aggressive and nonclinic children. *Journal of Abnormal Child Psychology, 19,* 553–567.

Dadds, M. R., & Sanders, M. R. (1992). Family interaction and child psychopathology: A comparison of two observation strategies. *Journal of Child and Family Studies, 1,* 371–391.

Danforth, J. S., Barkley, R. A., & Stokes, T. F. (1991). Observations of parent-child interactions with hyperactive children: Research and clinical implications. *Clinical Psychology Review, 11,* 703–727.

DeGarmo, D. S., & Forgatch, M. S. (2012). A confidant support and problem solving model of divorced fathers' parenting. *American Journal of Community Psychology, 49,* 258–269.

Dishion, T. J., Shaw, D., Connell, A., Gardner, F., Weaver, C., & Wilson, M. (2008). The family check-up with high risk families: Preventing problem behavior by increasing parents' positive support in early childhood. *Child Development, 29,* 1395–1414.

Eyberg, S. M., & Pincus, D. (1999). *Eyberg child behavior inventory and sutter-eyberg student behavior inventory-revised: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Fagot, B. I., & Gauvain, M. (1997). Mother–child problem solving: Continuity through the early childhood years. *Developmental Psychology, 33,* 480–488.

Fergusson, D. M., Lynskey, M. T., & Horwood, L. J. (1993). The effect of maternal depression on maternal ratings of child behavior. *Journal of Abnormal Child Psychology, 21,* 245–269.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). Implementation research: A synthesis of the literature. Tampa, FL: University of South Florida, Louis de la

Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

Gardner, F. (2000). Methodological issues in the direct observation of parent–child interaction: Do observational findings reflect the natural behavior of participants? *Clinical Child and Family Psychology Review, 3,* 185–198.

Gardner, F. E. M., Sonuga-Barke, E. J. S., & Sayal, K. (1999). Parents anticipating misbehavior: An observational study of strategies parents use to prevent conflict with behaviour problem children. *Journal of Child Psychology and Psychiatry, 40,* 1185–1196.

Gardner, F., Burton, J., & Klimes, I. (2006). Randomized controlled trial of a parenting intervention in the voluntary sector for reducing child conduct problems: Outcomes and mechanisms of change. *Journal of Child Psychology and Psychiatry, 47,* 1123–1132.

Gardner, F., Ward, S., Burton, J., & Wilson, C. (2003). The role of mother–child joint play in the early development of children's conduct problems: A longitudinal observational study. *Social Development, 12,* 361–378.

Glasgow, R. E., Vogt, T. M., & Boles, S. M. (1999). Evaluating the public health impact of health promotion interventions: The RE-AIM framework. *American Journal of Public Health, 89,* 1322–1327.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38,* 581–586.

Goodman, R. (1999). The extended version of the strengths and difficulties questionnaire as a guide to child psychiatric caseness and consequent burden. *Journal of Child Psychology and Psychiatry, 40,* 791–799.

Goodman, R., & Scott, S. (1999). Comparing the strengths and difficulties questionnaire and the child behavior checklist: Is small beautiful? *Journal of Abnormal Child Psychology, 27,* 17–24.

Gorin, A. A., & Stone, A. A. (2001). Recall biases and cognitive errors in retrospective self-reports: A call for momentary assessments. In A. Baum, T. Revenson, & J. Singer (Eds.), *Handbook of health psychology* (pp. 405–413). Mahwah, NJ: Erlbaum.

Granic, I., & Lamey, A. V. (2002). Combining dynamic systems and multivariate analyses to compare the mother–child interactions of externalizing subtypes. *Journal of Abnormal Child Psychology, 30,* 265–283.

Hawes, D. J., & Dadds, M. R. (2006). Assessing parenting practices through parent-report and direct observation during parent-training. *Journal of Child and Family Studies, 15,* 555–568.

Kerig, P. K. (2001). Introduction and overview: Conceptual issues in family observational research. In P. K. Kerig & K. M. Lindahl (Eds.), *Family observational coding systems: Resources for systemic research* (pp. 1–22). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the depression anxiety stress scales* (2nd ed.). Sydney, NSW, AU: Psychology Foundation of Australia.

Metzler, C. W., Biglan, A., Embry, D. D., Sprague, J. R., Boles, S. M., & Kavanagh, K. A. (2008). Improving the well-being of adolescents in Oregon. Eugene, OR: Center on Early Adolescence, Oregon Research Institute.

Metzler, C. W., Sanders, M. A., Rusby, J. C., & Crowley, R. (2012). Using consumer preference information to increase the reach and impact of media-based parenting interventions in a public health approach to parenting support. *Behavior Therapy, 43,* 257–270.

Morawska, A., & Sanders, M. R. (2010). *The child adjustment and parent efficacy scale*. Brisbane, QLD, AU: Parenting and Family Support Centre.

Mrazek, P., Biglan, A., & Hawkins, J. D. (2003). Community monitoring systems: Tracking and improving the well-being of America's children and adolescents. Falls Church, VA: Society for Prevention Research.

Müller, J. M., Achtergarde, S., & Furniss, T. (2011). The influence of maternal psychopathology on ratings of child psychiatric symptoms: An SEM analysis on cross-informant agreement. *European Child & Adolescent Psychiatry, 20,* 241–252.

Nexus Management Consulting. (2011). *Summary report: Evaluation of the implementation of Triple P in NSW*. Retrieved from Families NSW website: www.families.nsw.gov.au/assets/triplep_eval_report.pdf

Patterson, G. R., Reid, J. B., & Dishion, T. J. (1992). *Antisocial boys: Vol. 4. A social interactional approach*. Eugene, OR: Castalia Publishing Co.

Prinz, R. J., Sanders, M. R., Shapiro, C. J., Whitaker, D. J., & Lutzker, J. R. (2009). Population-based prevention of child maltreatment: The US Triple P system population trial. *Prevention Science, 10,* 1–12.

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Health Services Research, 38*(2), 65–76.

Putnam-Hornstein, E., Webster, D., Needell, B., & Magruder, J. (2011). A public health approach to child maltreatment surveillance. *Child Abuse Review, 20,* 256–273.

Saldana, L., Chamberlain, P., Wang, W., & Hendricks-Brown, C. (2011). Predicting program start-up using the stages of implementation measure. *Administration and Policy in Mental Health and Mental Health Services Research, 39*(6), 419–425.

Sanders, M. R. (2010). Adopting a public health approach to the delivery of evidence-based parenting interventions. *Canadian Psychology, 51,* 17–23.

Sanders, M. R., Markie-Dadds, C., Rinaldis, M., Firman, D., & Baig, N. (2007). Using household survey data to inform policy decisions regarding the delivery of evidenced-based parenting interventions. *Child: Care, Health and Development, 33,* 768–783.

Sanders, M. R., Markie-Dadds, C., & Turner, K. M. T. (1996). *Positive parenting*. Brisbane, QLD, AU: Families International. Retrieved from UQ Parenting and Family Support Centre website: http://www.pfsc.uq.edu.au/research/publications/resources.html.

Sanders, M. R., Markie-Dadds, C., & Turner, K. M. T. (1998). *Practitioner's manual for Enhanced Triple P*. Brisbane, QLD, AU: Families International. Retrieved from UQ Parenting and Family Support Centre website: http://www.pfsc.uq.edu.au/research/publications/resources.html.

Sanders, M. R., & Morawska, A. (2010). *The parenting and family adjustment scale*. Brisbane, QLD, AU: Parenting and Family Support Centre.

Sanders, M. R., Ralph, A., Sofronoff, K., Gardiner, P., Thompson, R., Dwyer, S., & Bidwell, K. (2008). Every family: A population approach to reducing behavioral and emotional problems in children making the transition to school. *Journal of Primary Prevention, 29,* 197–222.

Schwarz, N., & Sudman, S. (Eds.). (1994). *Autobiographical memory and the validity of retrospective reports*. New York, NY: Springer.

Scott, S., Briskman, J., & Dadds, M. R. (2011). Measuring parenting in community and public health research using brief child and parent reports. *Journal of Child and Family Studies, 20,* 343–352.

Shelton, K. K., Frick, P. J., & Wootton, J. (1996). The assessment of parenting practices in families of elementary school-aged children. *Journal of Clinical Child Psychology, 25,* 317–327.

Snyder, J., & Stoolmiller, M. (2002). Reinforcement and coercion mechanisms in the development of antisocial behavior: The family. In J. B. Reid & G. R. Patterson (Eds.), *Antisocial behavior in children and adolescents: A developmental analysis and model for intervention* (pp. 65–100). Washington, DC: American Psychological Association.

Spoth, R., Guyll, M., Redmond, C., Greenberg, M., & Feinberg, M. (2011). Six-year sustainability of evidence-based intervention implementation quality by community-university partnerships: The PROSPER Study. *American Journal of Community Psychology, 48,* 412–425.

Stoolmiller, M., Eddy, J. M., & Reid, J. B. (2000). Detecting and describing preventive intervention effects in a universal school-based randomized trial targeting delinquent and violent behavior. *Journal of Consulting and Clinical Psychology, 68,* 296–306.

Stoolmiller, M., & Snyder, J. (2004). A multilevel analysis of parental discipline and child antisocial behavior. *Behavior Therapy, 35,* 365–402.

Webster-Stratton, C. (1985). Comparisons of behavior transactions between conduct-disordered children and their mothers in the clinic and at home. *Journal of Abnormal Child Psychology, 13,* 169–184.

Webster-Stratton, C., & Hammond, M. (1997). Treating children with early-onset conduct problems: A comparison of child and parent training interventions. *Journal of Consulting and Clinical Psychology, 65,* 93–109.

Wells, K. C., Chi, T. C., Hinshaw, S. P., Epstein, J. N., Pfiffner, L., Nebel-Schwalm, M., Owens, E. B., Arnold, L. E., Abikoff, H. B., Conners, C. K., Elliott, G. R., Greenhill, L. L., Hechtman, L., Hoza, B., Jensen, P. S., March, J., Newcorn, J. H., Pelham, W. E., Severe, J. B., Swanson, J., Vitiello, B., & Wigal, T. (2006). Treatment-related changes in objectively measured parenting behaviors in the multimodal treatment study of children with attention-deficit/hyperactivity disorder. *Journal of Consulting and Clinical Psychology, 74,* 649–657.

Zaslow, M. J., Weinfield, N. S., Gallagher, M., Hari, E. C., Ogawa, J. R., Egeland, B., Tabors, P. O., & De Temple. J. M. (2006). Longitudinal prediction of child outcomes from differing measures of parenting in a low-income sample. *Developmental Psychology, 42,* 27–37.

Chapter 13
Multiple Comparisons and Truncation Bias in Family Policy Research: Strategies from the Building Strong Families Evaluation

Quinn Moore and Robert G. Wood

The evaluation of programs and policies focused on the family presents a number of challenges. Family programs often aim to influence a wide range of interrelated outcomes that are difficult to measure, relevant for different levels and types of family relationships, and evolving over time. Therefore, evaluations of these programs must be thorough in capturing the range of targeted outcomes using well-conceived measures. However, the evaluation design must also be efficient in assessing effectiveness without overburdening study participants and while being mindful that the probability of spuriously identifying impacts as statistically significant increases with the number of outcomes examined.

A recently conducted evaluation of Building Strong Families (BSF), a program intended to improve child well-being by strengthening the relationships of unmarried new parents, provides an example of some of these challenges and the strategies to overcome them (Wood et al. 2010, 2012). BSF is a multifaceted program intended to influence a range of intermediate and longer term outcomes, including the quality of relationships between parents, father involvement, parenting, and ultimately, child economic well-being and socioemotional development. The evaluation of BSF was based on a large-scale, multisite implementation of the program with random assignment of program applicants to a treatment group that was offered program services or a control group that was not.

This chapter uses the BSF evaluation as an example to explore two challenges of particular relevance to family research: (1) managing multiple comparisons when evaluating a program that aims to influence a wide range of outcomes and (2) dealing with potential bias introduced by examining measures that are not defined for all sample members.

Q. Moore (✉) · R. G. Wood
Mathematica Policy Research, Princeton, NJ, USA
e-mail: QMoore@mathematica-mpr.com

R. G. Wood
e-mail: RWood@mathematica-mpr.com

S. M. McHale et al. (eds.), *Emerging Methods in Family Research,*
National Symposium on Family Issues 4, DOI 10.1007/978-3-319-01562-0_13,

An Overview of BSF and Its Evaluation

Children raised outside of two-parent households are at increased risk for a variety of negative health, academic, and behavioral outcomes (Amato 2001, 2005; Brown 2004; McLanahan and Sandefur 1994). In response to these concerns, BSF was launched in 2002 by the Office of Planning, Research, and Evaluation, Administration for Children and Families (ACF), U.S. Department of Health and Human Services. BSF developed, implemented, and tested voluntary programs that aimed to create stable and healthy homes for children by strengthening the couple relationships of unmarried new parents. Mathematica Policy Research conducted the evaluation of BSF under contract to ACF. Sample intake for this evaluation took place from July 2005 through March 2008.

The decision to launch an intervention focused on unmarried couples was motivated by findings from the Fragile Families and Child Well-being Study (http://www.fragilefamilies.princeton.edu). Among unmarried parents in the Fragile Families study, most were romantically involved, had supportive and affectionate relationships, and were hopeful about their futures together (Carlson et al. 2005). However, these hopes were unrealized for many of the couples. Within 5 years, more than 60 % were no longer in a romantic relationship and only 16 % of them were married (Center for Research on Child Well-Being 2007). BSF was therefore designed to intervene in the relationships of new unmarried parents, at a time when they were optimistic about the relationship's future, and to provide support in meeting the unique challenges of these couples.

In order to be eligible for BSF, couples had to meet the following criteria: (1) both members of the couple agreed to participate, (2) the couple was romantically involved, (3) the couple was expecting a baby together or had a baby younger than three months old, (4) the couple was unmarried at the time the baby was conceived, and (5) both members of the couple were at least 18 years old.[1]

All BSF programs had three components: (1) group sessions on relationship skills, (2) individual support from family coordinators, and (3) assessment and referral to support services. Individual programs also had some flexibility in the implementation of BSF, including the specific curriculum used in the group education component and how they provided support and referral services. The sponsoring agency for each program chose for their group education sessions one of three curricula developed specifically for BSF. The programs covered key topics such as communication, conflict management, and marriage, and involved 30–42 h of instruction. At each site, the BSF family coordinator was responsible for assessing family members' needs and referring them to appropriate services. In addition, the family coordinator reinforced relationship skills, provided emotional support, and encouraged participation in the

[1] The members of the couple had to be able to speak one of the languages in which BSF was offered in their location. All locations offered BSF in English, and some also offered BSF in Spanish. At intake, couples were screened for intimate partner violence. If there was evidence of intimate partner violence that could be aggravated by participating in BSF, the couple was not considered eligible for BSF and was referred to other services.

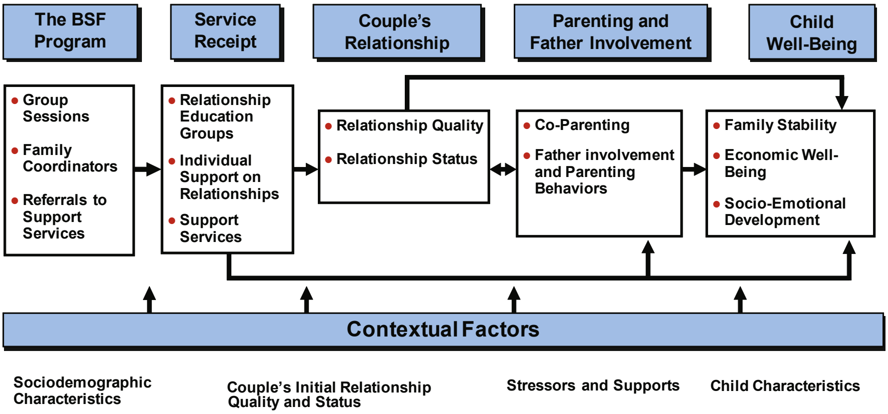


Fig. 13.1 Model of BSF and its expected impacts

group sessions. The support services could be provided by either the sponsoring organization or another provider in the community. Services could include those to address family members' housing problems, employment needs, or other issues.

Eight organizations in diverse locations around the United States were chosen by project and ACF staff to implement the BSF program. The programs were located in large metropolitan areas such as Atlanta, Georgia, and Houston, Texas, as well as smaller towns and cities such as San Angelo, Texas, and Baton Rouge, Louisiana. Most sponsor organizations developed BSF from the infrastructure of existing programs. Four local programs (located in Houston, San Angelo, Florida, and Indiana) added BSF services to their Healthy Families programs. Healthy Families programs aim to promote positive parenting and child development and to prevent child abuse and neglect via staff visitation and education of parents in their homes.

Mathematica Policy Research conducted a rigorous experimental evaluation of the eight BSF programs. The evaluation included more than 5,000 couples who applied and were eligible for BSF and were subsequently randomly assigned to either a group that could participate in BSF or a control group that could not. At the time of random assignment, Mathematica collected baseline information on members of both groups. Fifteen months after random assignment, telephone surveys collected data on how members of both research groups were faring. A follow-up took place close to the time of the child's third birthday. In addition to the telephone surveys, this follow-up included in-person assessments of the quality of the parent-child relationship and of the cognitive and socioemotional development of the children.

A model of how BSF could affect couples and their children guided the study design. The BSF intervention had the potential to affect multiple aspects of the lives of participating couples and their children. This evaluation examined the program's effects on outcomes within three broad areas: (1) the couple relationship, (2) parenting, and (3) child well-being (Fig. 13.1). Two outcome domains measured the couple relationship: (1) relationship quality and (2) relationship status. Two are associated with parenting: (1) the quality of the co-parenting relationship and (2) fathers' involvement and parenting behavior. Three are associated with child well-being: (1) children's family stability, (2) their economic well-being, and (3) their socioemotional development. We estimated program effects on outcomes in these domains by comparing the outcomes of the couples and families in the BSF group with the outcomes of those in the control group.

Analysis of BSF's impact 15 months after couples applied for the program focused on outcomes related to the couple relationship; at that point the focal children were not yet old enough to measure parenting and child well-being outcomes. This early follow-up analysis found that when data for the eight programs were combined, BSF had no effect on couples' relationship quality or the likelihood that they remained romantically involved or got married (Wood et al. 2010, 2011). However, the results varied across the eight programs included in the evaluation. The BSF program in Oklahoma City had a consistent pattern of positive effects on relationship outcomes, whereas the Baltimore program had a number of negative effects. The other BSF programs generally had little or no effect on relationships.

Analysis of a three-year follow-up survey also indicates that BSF did not succeed in its central objectives of improving the couple relationship, increasing the quality of coparenting, or enhancing father involvement (Wood et al. 2012). In fact, the program had modest negative effects on some of these outcomes. BSF also had little impact on child well-being, with no effect on children's family stability or economic well-being and only a modest positive effect on children's socio-emotional development. The impacts observed for the BSF programs in Oklahoma City and Baltimore at 15-month follow-up generally faded by the longer term follow-up, although the Oklahoma Cith program did increase the likelihood that children lived with both their biological parents until age three. At the three-year follow-up, the BSF program in Florida had negative impacts on relationship status and quality, co-parenting, father involvement, and family stability.

Strategies for Managing Multiple Comparisons

In BSF, as in other evaluations of family support policies, there are numerous potential outcomes of interest. As described above, researchers may be interested in multiple domains. Additionally, there may be uncertainty about how best to measure each domain of interest, leading to multiple dimensions or measures within a single domain. Researchers may also be concerned about the potential for variation in outcomes by subgroups, particularly if there is reason to believe that programs are succeeding in their goals for some populations but not for others. These research interests can lead to a long list of hypotheses to be tested.

However, interest in multiple hypotheses must be balanced against the concern that evaluating multiple hypotheses raises the risk of finding a statistically significant result by chance (Schochet 2009). For example, if 100 independent statistical tests are performed, with 5 % set as the threshold for statistical significance, on average, five results will be statistically significant by chance alone. Furthermore, this scenario has a 99 % likelihood of at least one statistically significant result that is due to chance. In policy research, the success or failure of programs is often determined by the statistically significant impacts produced. Therefore, it is especially important to mitigate the risk of spurious significant findings.

The central strategy for dealing with multiple comparisons in the BSF evaluation was to limit the number of outcomes that provide the key test of the effectiveness of BSF. By selecting primary outcomes before beginning the analysis and documenting the selections in an analysis plan, we avoid the temptation to focus post hoc on the outcomes that emerge as statistically significant while giving less attention to null findings.

Deciding which outcomes to prioritize requires a careful consideration of the program's goals. Key outcomes should be not only substantively important but also focused on areas in which the program is likely to have an impact. In BSF, we identified key outcomes based on the stated goals of the BSF program and the content of the curricula used. As noted earlier, in the 36-month follow-up analysis, the assessment of program effectiveness was based on the seven key domains BSF aimed most directly to affect: relationship status, relationship quality, co-parenting, father's involvement, family stability, economic well-being, and child socioemotional development (Fig. 13.1).

The analysis also examined BSF's effects on outcomes in several additional domains, such as attitudes toward marriage, mother's parenting behavior, and child language development. However, these domains were not primary areas of focus for the BSF program. Therefore, we include these and other outcomes as secondary results, which may provide suggestive evidence but do not provide the primary test of BSF's effectiveness.

Many of the key domains in the BSF evaluation included multiple measures. For example, the status of the couple's relationship was measured with three different binary outcomes: whether the couple was romantically involved, whether they were living together, and whether they were married. In total, we considered 20 outcomes within our seven key domains, as shown in Table 13.1. If BSF had no impact on any of the key outcomes, and all 20 impacts were independent, this would generate a 64 % chance of finding at least one statistically significant impact by chance.

In order to minimize the chance of overinterpreting statistically significant findings, the research team considered whether statistically significant impacts appeared to be isolated findings or were part of a robust pattern within the domain. One recommended strategy to formally test for the robustness of impacts within a domain is to generate composite measures for each domain and test for significant impacts on those composite measures (Schochet 2009). In the BSF analysis, we constructed a single composite measure for outcomes within each domain. For domains that included at least one continuous outcome, composite indices were constructed as the sum of normalized primary outcome measures within each domain (relationship quality, father involvement, and child socioemotional development). For domains that included only binary measures, the composite measure is the sum of the primary outcomes (relationship status and economic well-being). Lastly, for domains that

Table 13.1 Primary outcomes in key domains

Key domain	Primary outcome
Couple relationship domains	
Relationship status	Still romantically involved
	Living together (married or unmarried)
	Married
Relationship quality	Relationship happiness scale
	Support and affection abbreviated scale
	Use of constructive conflict behaviors scale
	Avoidance of destructive conflict behaviors scale
	Neither member of the couple was unfaithful since random assignment
Parenting domains	
Co-parenting	Quality of co-parenting relationship scale
Father's involvement and parenting behavior	Father lives with child
	Father regularly spends time with child
	Father's engagement with child
	Mother reports that father provides substantial financial support for raising child
	Father's parental responsiveness (observed)
Child well-being domains	
Family stability	Both parents have lived with child since birth
Economic well-being	Family's monthly income below poverty threshold
	Family experienced difficulty meeting housing expenses during past year
	Family receiving TANF or food stamps
Child socioemotional development	Behavior problems index
	Emotional insecurity amid parental conflict

had only a single primary outcome, that outcome was also used as the composite measure (co-parenting and family stability). Essentially, this reduced the number of comparisons from 20 (the number of outcomes in key domains) to 7 (the number of key domains).

As an additional robustness check on the strength of patterns of statistical significance, we made use of conventional statistical adjustments for multiple comparisons within each domain. These adjustments target an overall significance level within a domain by setting more stringent thresholds (*p*-values) at which individual statistical tests are considered significant. We used the Benjamini-Hochberg method, which takes into account both the number of comparisons and the strength of impacts in order to determine the thresholds at which *p*-values are considered statistically significant (Benjamini and Hochberg 1995).

The 15-month follow-up analysis of relationship status and quality for the Oklahoma City program provides an illustrative example of this approach. As shown in Table 13.1, these two domains include eight primary outcomes. Using conventional statistical tests with no adjustments for the number of comparisons, one of three measures of relationship status is statistically significant at the 10 % level, and four measures of relationship quality are statistically significant at the 5 % level, while the fifth is statistically significant at the 10 % level (Table 13.2). However, when

Table 13.2 Results from BSF 15-month follow-up analysis, Oklahoma City

Outcome	Statistical significance	
	No adjustment for multiple comparisons	Adjusted for multiple comparisons
Relationship status		
Composite Index	◯	N/A
Romantically involved	+	◯
Living together, married or unmarried	◯	◯
Married	◯	◯
Relationship quality		
Composite index	++	N/A
Relationship happiness scale	+ + +	++
Support and affection full scale	++	+
Use of constructive conflict behaviors scale	+ + +	++
Avoidance of destructive conflict behaviors scale	++	+
Fidelity	+	+

Source: BSF 15-month follow-up surveys, conducted by Mathematica Policy Research. Impacts are adjusted using a pooled regression controlling for the couple's baseline relationship and demographic characteristics. Adjustment for multiple comparisons is made using the Benjamini-Hochberg method. N/A = No adjustment is made for the composite index because it is analyzed independently as a single variable within the outcome domain
◯ No statistically significant impact
+ + +/++/+ Statistically significant positive impact at the .01/.05/.10 level

we adjusted for multiple comparisons using the Benjamini-Hochberg method, none of the impacts on relationship status were statistically significant, whereas three of the impacts on relationship quality outcomes were statistically significant at the 5 % level. Moreover, the impact on the composite index for relationship status was not statistically significant, whereas the impact on the index for relationship quality was statistically significant at the 5 % level. We conclude from these results that there is strong evidence that the Oklahoma City program had a positive effect on relationship quality, but much weaker evidence that it had a positive effect on relationship status.

Multiple Comparison Issues in Subgroup Analysis

Another source of testing of large numbers of hypotheses is the desire to understand variation in impacts among subpopulations. At minimum, the same steps taken to address multiple comparisons in the full sample (pooled) analysis should also be taken in the analysis of subpopulations. Thus, it is always appropriate to adjust for within-group comparisons.

In the evaluation of the eight individual BSF sites, we used the same approach to multiple comparisons used in the pooled analysis. Each site had flexibility in the program model it employed and in its recruitment methods. Thus, we treated each

Table 13.3 Subgroups examined in BSF impact analysis

Subgroup measure	Subgroup categories
Based on couple's initial relationship characteristics	
Initial relationship quality	Relationship quality index below the sample median
	Relationship quality index above the sample median
Multiple partner fertility	No children with other partners
	One or more children with other partners
Based on couple's sociodemographic and human capital characteristics	
Father's initial earnings	Father's total annual earnings $10,000 or less
	Father's total annual earnings greater than $10,000
Race/ethnicity: African American	Both partners are non-Hispanic African American
	All other couples
Race/ethnicity: Hispanic	Both partners are Hispanic
	All other couples
Young age	Either partner younger than age 21 at baseline
	Both partners age 21 or older at baseline

site as if it were a mini-evaluation and employed the same strategies for managing multiple comparisons as we did in the pooled analysis. We did not adjust further for comparisons across sites.

In other cases, it was appropriate to further adjust for the number of subpopulations considered. This is the approach we took when testing whether BSF's impacts varied across demographic groups. Although we had no prespecified hypotheses about variation in BSF impacts by subgroup, we wished to test whether there were specific demographic groups for which BSF appeared to work better (or worse). To limit the total number of comparisons, we tested for significant impact variation on only our composite indices in the seven key domains. Furthermore, we limited the number of subgroups by considering characteristics that capture key aspects of couples' relationships at baseline (relationship quality and multiple partner fertility) and key sociodemographic traits (race, ethnicity, age, and father's earnings). The subgroups considered are shown in Table 13.3. In addition to being substantively important, these characteristics include the subgroups for which there was some evidence of significant impacts 15 months after random assignment (race/ethnicity, age, education, initial relationship quality, and multiple partner fertility).

Even with our brief lists of outcomes and subgroups, there would be a substantial risk of reporting spurious findings if no additional steps were taken. If separate statistical tests are conducted for each of the 12 subgroup categories on the seven composite indices for the key domains, there are a total of 84 comparisons. Under the assumptions described above, this leads to a 99 % chance of finding at least one statistically significant difference, even if no true differences by subgroups exist. Rather than further limiting the outcomes or subgroups, we adopted a more stringent standard for featuring the subgroup-specific results in our main report. The purpose of this standard is to identify strong patterns of subgroup differences and reduce the probability of focusing on findings that are statistically significant because of chance.

First, we tested for statistically significant *differences* between the impact estimates for subgroup categories rather than focusing on the statistical significance of the subgroup impact estimates themselves. For example, this approach would assess whether the impact on a composite index for African American couples was significantly different than the impact for other couples, regardless of whether either of the two impacts is statistically significant on its own.

In addition to cutting the number of comparisons in half (in our example, from 12 individual subgroup categories to 6 subgroup differences), an approach focused on differences in impacts tracks more closely to the goal of subgroup analysis—to test whether the program works better (or worse) for some groups—than does an approach focused on individual subgroup impacts. For example, we would still be interested in a subgroup pattern that showed much stronger (or much weaker) impacts for African American couples than for other couples, even if impacts were not significant for either group. To further reduce the probability of focusing on subgroup findings that are due to chance, we decided, prior to conducting any analyses, that we would feature subgroup differences in impacts only if there was evidence (at the conventional 5 % significance level) of significant variation between subgroups in at least three of the seven domains.

The motivation for this rule is to call attention to variation in impacts by subgroup only if the variation appears to be robust across domains. Using our preferred rule requiring three statistically significant domain differences, we would have only a 14 % chance of identifying variation by subgroups as statistically significant even if significance is due only to chance. However, in developing the rule, we were mindful that an overly strict standard may preclude presenting meaningful subgroup findings. A more restrictive rule would provide a much lower probability of presenting a false positive but would also increase the chance that we would not be able to identify important subgroup findings.

Strategies for Dealing with Truncated Variables

Another common challenge in family policy research, as well as social science research more broadly, is that outcomes of interest can determine the population that is analyzed, which affects both the generalizability of research and the interpretation of comparisons based on such outcomes. Examples of this phenomenon—often referred to as truncation—include that academic performance is observed only for those who remain in school, earnings are observed only for those who are employed, and satisfaction with medical care is not observed for those who die (McConnell et al. 2008; Zhang and Rubin 2003). Because family research is often focused on outcomes that are determined by the existence of family relationships (marital quality, parenting practices, etc.), truncation is common.

Truncation has particularly strong implications for program evaluations. An evaluation using a random assignment design generally provides unbiased estimates of program effectiveness because the initial characteristics of the research groups can be expected to be equivalent, making any eventual differences between the outcomes attributable to the program. However, experimental estimates can be biased when the full sample that was randomly assigned is not included in the impact analysis. If an evaluated program determines the types of people for whom a truncated outcome is defined, then the initial characteristics of the treatment and control group members *for whom the outcome is available* may differ, leading to biased estimates of the program's effectiveness.

To see the potential for truncation bias in the BSF context, consider a scenario in which BSF has no effect on relationship quality but increases the number of couples who remain romantically involved by persuading couples with lower initial relationship quality to stay together. In this case, the estimated impact of BSF on relationship quality would be negative simply because the set of romantically involved couples (for whom relationship quality measures are defined) includes a larger number couples with lower relationship quality for the BSF group than for the control group. Because the true impact in this example is zero and the estimated impact is negative, the estimated impact is biased negatively because of truncation.

One simple strategy for dealing with truncation is to define a new measure in such a way that all sample members can be included in the analysis (McConnell et al. 2008). This is an effective strategy for avoiding truncation bias with truncated binary outcomes or truncated continuous outcomes that have meaningful cutoffs or thresholds. For example, evaluations of pregnancy prevention interventions may examine whether a person is engaging in unprotected sex (which is defined for everyone) rather than whether the person is using contraceptives (which is defined only for the sexually active).

Although this approach leads to rigorous impacts because it preserves the random assignment design, it is less satisfactory with truncated continuous variables that do not have natural thresholds. Consider a conflict management scale that is defined only for couples who interact regularly. It would be possible to construct a measure of whether the couple was in a relationship with "good" conflict management skills, that is: (a) equal to 1 if the couple interacts regularly and has a scale score above a certain threshold; or (b) equal to 0 if the couple has a scale score below the threshold or no longer interacts regularly. Although this variable would not be truncated, impact estimates may be sensitive to the threshold value and the choice of threshold could be criticized as arbitrary. Moreover, this binary outcome may ignore much of the richness and variation that the continuous scale score provides.

For truncated variables such as these, our strategy was to examine the truncated variable but to make a careful assessment of the risk of truncation bias. To make this assessment, we treated truncation as a type of sample attrition because, as with attrition, the measure is not available for the affected sample members. We then used established guidelines related to acceptable sample attrition in experimental evaluations. Based on the results of this risk assessment, the presentation of findings includes cautions for any analysis that has a moderate risk of bias. Analysis that has a high risk of bias was not included as part of the evaluation of program effectiveness and was only presented in a technical appendix with strong cautions to the reader.

At its core, the BSF evaluation's approach to outcome truncation is to assess the validity of the simple treatment-control group comparison offered by the experimental framework. An advantage of this approach is that it offers impact estimates that are easy to understand and that are consistent with the presentation of non-truncated analysis. This approach works for BSF because the key truncated analysis samples are at low or moderate risk of bias. This approach may not be satisfactory in studies that have truncated samples with substantial risk of bias. For such studies, quasi-experimental approaches that attempt to model selection into the truncated sample may be more appropriate.

Assessing Risk of Truncation Bias

The process for assessing the risk of bias due to truncation followed a two-step procedure developed for the U.S. Department of Education's What Works Clearinghouse (WWC; U.S. Department of Education 2008).

Step 1: Attrition Testing. First, the team analyzed whether each analysis sample met an attrition standard based on a combination of overall sample attrition and differential attrition between the BSF and control groups. The acceptable amount of one type of attrition depends on the amount of the other type. For instance, the WWC *Procedures and Standards Handbook* (U.S. Department of Education 2008) notes that "bias associated with an overall attrition rate of 10 % and a differential attrition rate of 5 % can be equal to the bias associated with an overall attrition rate of 30 % and a differential attrition rate of 2 %" (p. 28).

The WWC has developed two sets of attrition standards based on the extent to which attrition is likely to be associated with the intervention under study. In cases for which much of the attrition is likely to be exogenous, the more liberal standards may be used. In cases for which attrition is likely to be related to the intervention, the more conservative standards should be used (U.S. Department of Education 2008). For the BSF analysis, both instrument nonresponse and item truncation are treated as attrition. Because BSF may affect whether couples remain in contact or romantically involved, attrition due to truncation may be related to program participation. Therefore, we used the conservative WWC attrition standards in assessing the risk of attrition bias. If the conservative attrition standard is met, then the analysis meets WWC standards and the risk of serious bias due to attrition is deemed low.

Step 2: Initial Equivalence Testing. If a sample used for an impact analysis failed to meet the WWC attrition standard, then the evaluation team proceeded to the second step and tested the two research groups for equivalence on observable characteristics. In order to meet WWC initial equivalence standards, the effect size of the difference between treatment and control group means for each of the key baseline measures must be less than one quarter of a standard deviation.

For analysis samples using follow-up survey data, we used the following baseline variables for equivalence testing: (1) a relationship commitment scale, (2) a relationship interaction scale, (3) relationship marital and cohabitation status, and (4) race/ethnicity. The evaluation team selected the baseline measures of relationship quality and status for these tests because measures in these domains represent

Table 13.4 Truncated BSF impact analysis samples and associated primary outcomes

Analysis sample	Primary outcomes examined using these samples
Either parent responded and couple in regular contact	Relationship quality (*constructive and destructive conflict behaviors*) Child socioemotional development (*emotional insecurity amid parental conflict*)
Either parent responded and couple romantically involved	Relationship quality (*relationship happiness, support and affection*)

the outcomes of most central importance for the impact analysis. Measures of race/ethnicity were included because of large differences in marital status, relationship dissolution, and relationship quality between racial and ethnic groups documented in prior literature (Brown 2003; Graefe and Lichter 2002; Tucker and Mitchell-Kernan 1995). In addition, race/ethnicity and initial relationship status and quality are the baseline measures that are the most highly predictive of the key relationship outcomes examined in the analysis.

For equivalence testing for analyses based on direct assessments, we selected measures that reflect the direct assessment's focus on child development and parent interactions with the child and included: (1) child's gender, (2) race/ethnicity, (3) respondent's age, and (4) pregnancy intendedness.

Applying WWC Evidence Standards to the BSF 36-Month Analysis

The 36-month impact analysis examines outcomes in two truncated analysis samples: couples in regular contact and couples who were still romantically involved (Table 13.4). We applied the WWC evidence standards to both truncated analysis samples (as well as to nontruncated analysis samples) and conducted these tests for both the pooled samples and the samples for each of the eight local programs.

For pooled analyses, the sample of couples in regular contact met WWC evidence standards (Table 13.5). Therefore, analysis based on this sample was determined to have low risk of attrition bias. The sample of romantically involved couples did not meet the WWC's conservative attrition standards. However, the BSF and control groups had differences in examined baseline characteristics of less than a quarter of a standard deviation. Therefore, this analysis meets WWC evidence standards with reservations and is determined to have moderate risk of attrition bias. This analysis is presented as a part of the evaluation of BSF's effectiveness, but the report alerts readers that these results should be interpreted more cautiously than other experimental impacts should.

Most site-level analysis samples of couples in regular contact meet WWC evidence standards with or without reservations. However, most site-level analysis samples of romantically involved couples do not. As noted above, findings that do not meet standards are not included in the assessment of program effectiveness.

Table 13.5 Results of assessments of risk of attrition bias for truncated BSF analysis samples

	Low attrition standard met?	Initial equivalence standard met?	WWC rating
Samples of couples who were in regular contact at 36-month follow-up			
Pooled across programs	Yes	N/A	Meets standards
Atlanta	No	Yes	Meets standards with reservations
Baltimore	Yes	N/A	Meets Standards
Baton Rouge	No	Yes	Meets standards with reservations
Florida counties	No	Yes	Meets standards with reservations
Houston	No	No	Does not meet standards
Indiana counties	Yes	N/A	Meets standards
Oklahoma City	Yes	N/A	Meets standards
San Angelo	Yes	N/A	Meets standards
Samples of couples who were romantically involved at 36-month follow-up			
Pooled across programs	No	Yes	Meets standards with reservations
Atlanta	No	Yes	Meets standards with reservations
Baltimore	No	No	Does not meet standards
Baton Rouge	No	No	Does not meet standards
Florida counties	No	No	Does not meet standards
Houston	No	No	Does not meet standards
Indiana counties	No	No	Does not meet standards
Oklahoma city	Yes	N/A	Meets standards
San Angelo	No	No	Does not meet standards

Source: BSF 36-month follow-up surveys and 36-month direct assessments, conducted by Mathematica Policy Research. Attrition is determined by both survey nonresponse and measure truncation. A couple is counted as responding if at least one partner completed a 36-month survey. For the relationship happiness and support and affection measures, the rate of truncation is the percentage of responding couples who are no longer in a romantic relationship at 36-month follow-up. For the two conflict behavior scales, the truncation rate is the percentage of responding couples who are no longer in regular communication. Analysis samples that meet WWC standards with reservations are determined to have moderate risk of attrition bias. Analysis samples that do not meet WWC standards are determined to have substantial risk of attrition bias. Findings related to these samples are presented only in appendices to the technical supplement and not in the main report. N/A = Not applicable. Do not conduct initial equivalence test if the attrition standard is met

Implications

Multiple comparisons and truncation bias are issues that have important implications for the interpretation of research findings. Without accounting for the number of hypotheses being tested, researchers are at risk of overstating the prevalence of statistically significant findings (both positive and negative). Without considering the populations for which the analyzed outcomes are defined, researchers run the risk of misinterpreting findings or presenting biased results. The BSF evaluation provides a useful example (among many potential satisfactory approaches) of strategies for addressing these challenges.

The BSF evaluation's core strategy for addressing multiple comparison concerns is simple: It is to consider the central goals of the program carefully before beginning data collection or analysis and then use these goals to inform development of a compact list of outcomes that can capture program effectiveness. Developing a well formulated set of hypotheses and a set of outcomes that will allow them to be tested is at the heart of the scientific method. Developing these hypotheses before beginning analysis should be fundamental to any research endeavor, whether related to family policy studies or not.

Given the often ambitious goals of family programs and policy, it can be difficult to limit outcomes to a number for which the risk of spurious statistically significant findings is acceptably low. Thus, it is important for family researchers to be aware of other tools available for addressing multiple comparison concerns. These include various statistical adjustments for multiple comparisons and indices that summarize findings within an outcome domain. However, perhaps the most powerful tool in the researchers' arsenal is a clear, honest, and complete presentation of findings. If readers are informed of all hypotheses that were tested, as well as how and when these hypotheses were formulated, they will be much more equipped to assess which findings appear to be part of a strong pattern within a domain and which are more isolated.

While there is a need for a transparent process for identifying the outcomes to be used as primary assessments of program effectiveness, there is also an important place for secondary or exploratory analysis of additional outcomes or subpopulations. Exploratory analyses are particularly appropriate as a part of efforts to understand more deeply the important findings related to the primary analysis. However, these findings should not be used to draw conclusions about program effectiveness, and their exploratory nature should be clearly indicated to the reader.

Another benefit of a transparent and well articulated process for identifying the primary assessments of program effectiveness is that it helps defuse criticism that the evaluation may have been "rigged" to produce a particular result. Transparency is particularly valuable for family programs and policy, which are often politically charged. For example, some observers were opposed to healthy marriage programs such as BSF because they felt that relationship-focused interventions could only indirectly affect child outcomes and may have detrimental impacts if they encouraged couples with poor relationship quality to stay together. Other observers felt equally strongly that interventions that encourage relationship commitment and father involvement are vital to child well-being. A transparent evaluation process helped assure both constituencies that the program's effectiveness had been assessed fairly and objectively. Evaluations of other family programs, such as abstinence-only education, gain similar benefits from transparency.

A preferred strategy for dealing with truncated variables is to reformulate the truncated variable into a binary variable that is defined for all sample members (as in the example discussed above of contraceptive use in pregnancy prevention evaluations). However, this approach does not work well when the truncated variable is continuous and does not have meaningful threshold values, as is the case for relationship happiness, conflict management, and many other scales relevant to family research. The BSF evaluation's innovative strategy for dealing with truncated variables is to assess the risk of bias that truncation poses based on existing standards for sample attrition in experimental evaluations and then use that assessment to determine which analysis may be included in evaluating program effectiveness. Again, clear presentation is vital for readers to understand how truncation may or may not influence findings and which findings are more likely to be subject to bias. The BSF evaluation report informs readers when truncation may pose a moderate risk of bias and excludes analysis with a high risk of bias.

While the BSF evaluation provides an example of how truncation bias can affect experimental evaluations of family programs, it is important to emphasize that this issue is relevant to a much broader set of family research. Any comparisons using truncated outcomes require careful thought about how the conditioning variable should affect the interpretation of findings. For example, a study of racial difference in risk of divorce or inter-partner household production patterns must consider racial differences in marriage and factors related to propensity to marry. Similarly, a study examining gender differences in labor force outcomes must consider gender differences in employment and household responsibility. As with the BSF evaluation, each of these cases requires that the researcher assess the risk of truncation bias, develop a plan for addressing potential bias, and clearly inform readers of how truncation may influence findings.

References

Amato, P. (2001). Children of divorce in the 1990s: An update of the Amato and Keith meta-analysis. *Journal of Family Psychology, 15,* 355–370.

Amato, P. (2005). The impact of family formation change on the cognitive, social, and emotional well-being of the next generation. *Future of Children, 15*(2), 75–96.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological), 57,* 289–300.

Brown, S. L. (2003). Relationship quality dynamics of cohabiting unions. *Journal of Family Issues, 24*(5), 583–601.

Brown, S. L. (2004). Family structure and child well-being: The significance of parental cohabitation. *Journal of Marriage and Family, 66,* 351–367.

Carlson, M., McLanahan, S., England, P., & Devaney, B. (2005). What we know about unmarried parents: Implications for building strong families programs. http://www.mathematica-mpr.com/publications/redirect_PubsDB.asp?strSite=PDFs/bsfisbr3.pdf. Accessed Aug 2012.

Center for Research on Child Well-being. (2007). *Parents' relationship status five years after a non-marital birth* (Fragile Families Research Brief No. 39). Princeton: CRCW.

Graefe, D. R., & Lichter, D. T. (2002). Marriage among unwed mothers: Whites, blacks and Hispanics compared. *Perspectives on Sexual and Reproductive Health, 34*(6), 286–293.

McConnell, S., Stuart, E. A., & Devaney, B. (2008). The truncation by death problem. *Evaluation Review, 32*(2), 157–186.

McLanahan, S., & Sandefur, G. (1994). *Growing up with a single parent: What hurts, what helps*. Cambridge: Harvard University Press.

Schochet, P. Z. (2009). An approach for addressing the multiple testing problem in social policy impact evaluations. *Evaluation Review, 33*(6), 539–567.

Tucker, M. B., & Mitchell-Kernan, C. (1995). Trends in African American family formation: A theoretical and statistical overview. In M. B. Tucker, & C. Mitchell-Kernan (Eds.), *The decline in marriage among African Americans* (pp. 3–26). New York: Russell Sage.

U.S. Department of Education. (2008). *WWC procedures and standards handbook*. Washington, DC: Institute for Education Sciences. http://ies.ed.gov/ncee/wwc/references/idocviewer/Doc.aspx?docId=19&tocId=7. Accessed Aug 2012.

Wood, R. G., McConnell, S., Moore, Q., Clarkwest, A., & Hsueh, J. (2010). *Strengthening unmarried parent's relationships: The early impacts of building strong families*. Princeton: Mathematica Policy Research.

Wood, R. G., Moore, Q., & Clarkwest, A. (2011). BSF's effects on couples who attended group relationship skills sessions: A special analysis of 15-month data. Princeton: Mathematica Policy Research.

Wood, R. G., Moore, Q., Clarkwest, A., Killewald, A., & Monahan, S. (2012). *The long-lerm effects of Building Strong Families: A relationship skills education program for unmarried parents*. Princeton: Mathematica Policy Research.

Zhang, J. L., & Rubin, D. B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *Journal of Educational and Behavioral Statistics, 28*(4), 353–368.

Chapter 14
Optimizing Family Intervention Programs: The Multiphase Optimization Strategy (MOST)

Linda M. Collins

Optimizing Family Intervention Programs: The Multiphase Optimization Strategy (MOST)

Behavioral interventions aimed at the family are an important vehicle for improving the well-being of the family as a unit and of its individual members. A few examples include the Family Bereavement Program (Hagan et al. 2012; Sandler et al. 2010), aimed at helping to improve adjustment in parentally bereaved families; the Family Check-Up (Shaw et al. 2006), an intervention for reducing child problem behavior by improving parenting; and Familias Unidas (Prado and Pantin 2011), aimed at preventing substance use and risky sexual behavior in Hispanic youth.

As different as these behavioral interventions are, they have some important features in common. First, each of them is multi-component. The Family Bereavement Program includes separate components targeted to children, adolescents, and caretakers. The Family Check-Up includes components focusing on parenting, family management, and contextual issues. Components of Familias Unidas include family visits, group sessions with parents, parent-teacher/counselor meetings, skill building activities, and supervised peer activities.

The second feature that these behavioral interventions have in common is that they are all evidence-based; that is, each has been empirically evaluated and demonstrated to have a statistically significant effect on key outcomes. The randomized clinical trial (RCT) is widely considered to be the gold standard for this evaluation. The purpose of an RCT is to compare experimentally the performance of a behavioral intervention to that of a suitable comparison intervention. The comparison intervention is frequently whatever constitutes the customary standard of care, or sometimes treatment is delayed in the control group and provided after the intervention is expected to have its effect in the treatment group.

L. M. Collins (✉)
The Methodology Center, Department of Human Development and Family Studies,
The Pennsylvania State University, University Park, PA, USA
e-mail: lmc8@psu.edu

S. M. McHale et al. (eds.), *Emerging Methods in Family Research,*
National Symposium on Family Issues 4, DOI 10.1007/978-3-319-01562-0_14,

Although evaluation of multicomponent behavioral interventions via the RCT is indispensable, much more can be done to develop highly effective, efficient, and cost-effective behavioral interventions (Collins et al. 2005, 2007, 2011). As an alternative to sole reliance on the RCT, Collins and colleagues suggest a different approach called the multiphase optimization strategy (MOST). MOST is a comprehensive framework that includes not only evaluation of behavioral interventions via the RCT, but also optimization of the intervention before evaluation. This is done by conducting a component screening experiment to gather empirical information on the performance of individual intervention components, and using this information to select the intervention components and component levels to form an intervention package that meets a prespecified optimization criterion. In other words, the objective is to build an optimized intervention, and then to bring this optimized intervention to an RCT. Examples of applications of the MOST framework include Caldwell et al. (2012), Collins et al. (2011), McClure et al. (2012), and Strecher et al. (2008).

The purpose of this chapter is to introduce MOST to the family research community. The chapter will review a hypothetical example of the use of MOST to develop a behavioral intervention, and then offer a discussion of considerations relevant to the application of MOST in family research.

A Hypothetical Example of MOST

Why Consider Using MOST?

Suppose a team of investigators is interested in developing a behavioral intervention for treatment of adolescent substance use. The investigators are considering four components for inclusion in the intervention. The first three components are aimed at the adolescent. They are (1) motivational enhancement and cognitive behavioral therapy (ME/CBT; Sampl and Kadden 2001); (2) academic skill building (ASB), in which adolescents are taught study and time management skills, and tutored in their weakest school subjects; and (3) contingency management (CM; Stanger et al. 2009), in which the adolescent undergoes a weekly urine toxicology screening and is given a monetary reward for each clean specimen. The final component is (4) multidimensional family therapy (MDFT; Liddle et al. 2008), in which a therapist works with the adolescent and the family in several different treatment domains, including parenting skills, family communication, and family competence.

The customary approach would be to form an intervention package by combining the four components, and then evaluate this intervention by means of a classic RCT. This approach would address an important question: Does the intervention, as a package, have a statistically significant effect on adolescent substance use compared to a suitable control or comparison group? However, the RCT would not directly address some other questions, which are also important. Suppose the RCT is conducted, and the results suggest that the intervention package has a statistically significant effect. The RCT would still not provide direct answers to the following questions:

Does each of the four components contribute to the overall effect, and therefore, is each essential to the intervention? It may be that one or more of the components is not effective, or does not provide added value over and above other components. In this case these components could be eliminated to save time and money, with little or no decrement in overall effectiveness.

Does each component produce an effect large enough to justify its cost? For example, CM can be very resource-intensive. In Stanger et al.'s (2009) CM procedure each adolescent could earn up to nearly $600. In addition, CM can be somewhat controversial politically, suggesting that there can be costs in terms of time and effort associated with convincing communities to support its use. If CM is more expensive than other components that are candidates for inclusion in the intervention, it is worthwhile to determine whether it also has a larger effect.

Now suppose the RCT suggests that the intervention package does not have a statistically significant effect. The RCT would not provide direct answers to the following question: *Are any of the individual components worth saving for inclusion in a future intervention?* If an RCT shows that a treatment package is ineffective, the investigator learns little about what went wrong. It would be helpful to know whether one or more of the components is worth saving. It is even possible that one or more components had an iatrogenic effect, offsetting the positive effects of other components and producing a net effect of zero.

It is not necessary to address any of these questions to obtain evidence about whether the treatment as a package is working as intended. However, it is necessary to address these questions to develop an efficient intervention, point the way toward the next steps that should be taken to improve a successful intervention, or amend an unsuccessful intervention and make it successful. These are reasons why our hypothetical investigators might decide to use the MOST framework to develop an adolescent substance use treatment intervention.

Optimization and the Optimization Criterion

The words "optimal" and "optimized" are sometimes used colloquially to mean "best." However, in MOST a more technical definition drawn from mathematics and engineering is used. According to the *Concise Oxford Dictionary of Mathematics*, the term "optimization" means "the process of finding the best possible solution ... subject to given constraints" (Clapham and Nicholson 2009). Thus, according to this definition an optimized intervention is not the best possible intervention in some absolute sense. Rather, it is the best intervention that can be constructed subject to realistic constraints or limitations. In MOST, one limitation is always the finite set of intervention components under consideration. Other constraints may be the amount of money or the amount of time required to deliver the intervention.

MOST requires the investigator to choose an *optimization criterion*. The optimization criterion is an operational definition of an optimized intervention. In other words, it is the goal of MOST as applied to a particular intervention. In our example we will use one simple optimization criterion: “an intervention that includes only active components operating in the desired direction.” This criterion expresses the modest goal of arriving at an intervention that includes only components that have been demonstrated to contribute positively to any overall program effect, and is the optimization criterion that will be used in the hypothetical example described here. More sophisticated optimization criteria are possible. For example, suppose it is known that for pragmatic reasons the drug abuse treatment program must cost no more than $1,000 per family to implement. Then the desired optimization criterion might be “the most effective intervention (based on the four components at hand) that will not exceed $1,000 per family in implementation costs.” It is also possible to express the optimization criterion in terms of an absolute effect size, for example, to come as close as possible to an overall effect size of $d = 0.5$.

The optimization criterion should articulate a goal for the intervention that is clear, is desirable, and includes realistic constraints. Suppose the investigators decided to select as an optimization criterion “the most effective intervention that will not exceed $5,000 per family in implementation costs.” The resulting intervention would likely be more effective than an intervention developed using the constraint of $1,000 per family, but this means little if the $5,000 intervention is unlikely to be implemented broadly. It may be better to select an optimization criterion that uses a smaller upper limit on implementation costs and arrive at a somewhat less effective but readily implementable intervention. Of course, the resulting intervention must still demonstrate a statistically significant effect when it is evaluated via an RCT.

The Component Screening Experiment

The purpose of the component screening experiment is to gather information that will be used to determine which intervention components are eligible for inclusion in the intervention. (Please note that the term “screening” is used here the same way it is used in engineering: it refers to screening of components rather than screening of intervention participants based on characteristics such as psychiatric diagnosis.) Selection of an experimental design is critical. Taking a resource management perspective when selecting an experimental design is recommended (Collins et al. 2009). From this perspective, the best experimental design is the one that gathers the most, and most relevant, scientific information while making the most efficient use of available resources. The term “resources” is broadly defined here. Usually money and experimental subjects are the most important resources, but the list may include personnel, equipment, time, or any other resource that must be drawn upon in significant quantities to conduct the experiment.

In the example, there are four intervention components to be examined. One alternative would be to conduct four individual experiments, one for each of the intervention components. That is, one experiment would compare ME/CBT to a control, a second experiment would compare ASB to a control, and so on. In a sense,

Table 14.1 Experimental conditions in factorial component screening experiment

	Intervention component			
Condition number	ME/CBT	ASB	CM	MDFT
1	On	On	On	On
2	On	On	On	Off
3	On	On	Off	On
4	On	On	Off	Off
5	On	Off	On	On
6	On	Off	On	Off
7	On	Off	Off	On
8	On	Off	Off	Off
9	Off	On	On	On
10	Off	On	On	Off
11	Off	On	Off	On
12	Off	On	Off	Off
13	Off	Off	On	On
14	Off	Off	On	Off
15	Off	Off	Off	On
16	Off	Off	Off	Off

each of these experiments is a mini-RCT, in which a single treatment is compared to a control.

Another alternative is to conduct a factorial experiment (e.g., Fisher 1926; Kirk 2012). In a factorial experiment, multiple intervention components can be examined simultaneously, in a design that varies the levels of the factors in a systematic way. In the example, each of the intervention components can have two levels: on (included in the intervention) or off (not included in the intervention). Table 14.1 shows the experimental conditions that would be included in the factorial experiment. As Table 14.1 shows, in this factorial experiment every level of each intervention component is systematically combined with every level of every other intervention component. This results in a total of $2^4 = 16$ experimental conditions.

The logical underpinnings of a factorial experiment are different from those of an RCT and related experimental designs. In an RCT, individual experimental conditions are directly compared to assess the differences between them. By contrast, in a factorial experiment individual experimental conditions are usually not compared, and therefore statistical power is not based on comparisons of individual experimental conditions. Instead, combinations of experimental conditions are compared to estimate the main effect of each independent variable (e.g., intervention component), and the interactions between independent variables. For example, the main effect of ME/CBT would be obtained by comparing the mean of all of the experimental conditions in which ME/CBT is set to on (conditions 1–8) against the mean of all of the experimental conditions in which ME/CBT is set to off (conditions 9–16). Similarly, to obtain the main effect of ASB, the mean of the conditions in which ASB is set to on (conditions 1–4 and 9–12) would be compared to the mean of the conditions in which ASB is set to off (conditions 5–8 and 13–16). The main effects of the other two intervention components are obtained in the same manner. Note that all of the subjects are included in all comparisons by “reshuffling” the subjects for each comparison. The same subject might be in the “off” group for ME/CBT, the “on” group for ASB, and so on.

The reader may wonder how it can be appropriate to conduct this comparison when subjects in, say, the ME/CBT “on” group are in different combinations of levels of the other intervention components. The answer is that complete factorial experiments (as opposed to fractional factorial experiments, which are described later in this chapter) are designed to eliminate confounding of effects. Examination of Table 14.1 shows that when ME/CBT is set to on, each of the other intervention components is on in half of the experimental conditions and off in half of the experimental conditions, and this pattern is repeated when ME/CBT is set to off. Thus, although the other intervention components may have an effect on the outcome, if they raise the mean when ME/CBT is set to on, they will raise the mean an equivalent amount when ME/CBT is set to off, and there will be no net effect on the estimate of the main effect. The same holds true for estimates of interactions between components.

Selecting an Experimental Design from a Resource Management Perspective

Now that the logic of the factorial experiment has been reviewed, we can compare the scientific yield and resource demands of four individual experiments vs. one factorial experiment. Both approaches yield estimates of the individual effect of each component, although the estimates are not equivalent (see Collins et al. 2009). Only the factorial experiment yields estimates of interactions between components.

To compare resource demands it is necessary to consider the sample size requirements to achieve the desired level of statistical power. The investigators who are to conduct the screening experiment must identify the minimum effect size that a component must demonstrate to render it eligible for inclusion in the intervention package. This effect size then serves as the starting point for a power analysis, because effects smaller than this predetermined minimum are not of interest for the purpose of intervention optimization. Suppose the investigators have decided that an intervention component must demonstrate a main effect size of at least $d = 0.3$ to be considered for inclusion in the intervention package. A power analysis indicates that achieving power of at least 0.80 requires a sample size of approximately 340 for each experiment, or a total N of 1,360 for the individual experiments approach. By contrast, the power analysis indicates that in the factorial experiment, a total N of approximately 340 is required to detect the same effect size at the same level of statistical power. In other words, the individual experiments approach requires four times the sample size of the factorial experiment. The overall N in the factorial experiment will be divided approximately equally among the 16 experimental conditions, for a per-condition sample size of about 21. Although the factorial experiment has relatively few subjects per condition, this does not matter because, as mentioned above, individual conditions are not compared. Thus, the power of a factorial experiment is driven by the overall N rather than the number of subjects in individual conditions.

Although a factorial experiment always requires many fewer subjects than would be required to conduct individual experiments on the same number of intervention components, the trade-off is that it will nearly always require more experimental conditions. As Table 14.1 shows, the factorial experiment will require 16 experimental conditions, as opposed to a total of eight in the individual experiments approach. The additional experimental conditions required by the factorial experiment may result in additional overhead costs, such as additional staff to deal with the more complicated logistics.

In choosing between these two alternative design approaches and others, the investigators must weigh the benefits against the costs. The factorial experiment yields both main effects and interactions, and in this case requires one quarter as many experimental subjects. However, the factorial experiment also requires implementation of twice as many experimental conditions. The decision hinges on what resources are available, and which approach makes the best use of these resources. A SAS macro that investigators can use to weigh the relative resource demands of different designs for component screening experiments may be found at http://methodology.psu.edu/downloads.

Decision Making Based on Component Screening

The data from the screening experiment (if the factorial design is chosen) or experiments (if the individual experiments approach is chosen) are analyzed to produce estimates of the individual component effects and, in the case of a factorial experiment, interactions between intervention components. Based on the results, the investigators can determine which of the components attain the prespecified threshold effect size of $d = 0.3$. If interaction effects are available, these can be used to determine whether the effect of a component is significantly enhanced or dampened by the presence of one or more other components. This is all the information that is needed to achieve the simple optimization criterion of "an intervention that includes only active components." Any components that achieve or exceed the threshold effect size will be included in the intervention package. If a more complex optimization criterion has been selected, it may be necessary to include additional data, such as implementation costs, in decision making.

If the intervention package is satisfactory, the next step is to conduct a standard RCT, with individual families randomly assigned either to receive the intervention package or to a control group, comparison group, or wait list. However, it is possible that the results of the screening experiment will indicate that it is not worthwhile to devote resources to an RCT. This conclusion might be drawn if too few or even no components reach the threshold effect size. In this case the resources that would have been devoted to an RCT could instead be devoted to another screening experiment to examine a new set of candidate intervention components. Any components that were demonstrated to be successful in the previous screening experiment could either be included in the new experiment to replicate their effects and determine whether they interact with any of the new components, or set aside for inclusion in the treatment package when it is ultimately assembled.

Some Considerations Relevant to the Application of MOST in Family Research

What Constitutes an Intervention Component?

The concept of an intervention component is a broad one. In most cases, the investigator determines what constitutes a component, selecting aspects of the intervention that (1) can be separated out for examination and (2) are important enough to devote resources to examining. In the hypothetical example reviewed here, the intervention components were distinct strategies for adolescent drug abuse treatment. It would be possible to apply the MOST framework to examine a different set of intervention components. Four treatment domains within multidimensional family therapy (MDFT): adolescent, parent, interactional, and extrafamilial, are discussed by Liddle et al. (2008). If the treatments associated with each of these domains can be delivered independently, that is none requires the inclusion of any of the others, then MOST could be used to develop an optimized MDFT intervention. Sometimes there is a core component that provides a foundation upon which all of the others build, such as a component providing basic information about the effects of drugs in a drug abuse prevention or treatment program. An experimental design can be used in which this core component is provided to all subjects, as long as the remaining components are independent of each other. With this approach the effectiveness of the core component alone cannot be estimated.

The components examined in MOST do not necessarily have to be aspects of intervention content. A study in which the intervention itself, a school-based drug abuse and HIV prevention program focusing on positive use of leisure time, was not examined is described by Caldwell et al. (2012). Instead, the investigators examined three factors hypothesized to affect the fidelity with which teachers delivered the intervention: enhanced teacher training, teacher support, and enhanced school environment. It would also be possible to use MOST to examine components that promote adherence to the intervention, or that support sustainability of implementation.

Highly Efficient Fractional Factorial Experimental Designs

As was discussed in an earlier section, factorial experiments often require many fewer experimental subjects than alternative approaches. However, they usually require implementation of more experimental conditions. As the number of intervention components to be examined increases, the number of experimental conditions required rises rapidly (but not the number of subjects required; see Collins et al. 2009).

If another two-level intervention component were to be added to the hypothetical factorial experiment depicted in Table 14.1, the number of experimental conditions would be doubled, to $2^5 = 32$; examination of six components would require $2^6 = 64$ experimental conditions. Although the logistics of large factorial experiments in field settings should not be undertaken lightly, there are examples of such experiments being carried out successfully, even in challenging circumstances. One example provided by Caldwell et al. (2012) described an eight-condition factorial experiment in middle schools in a poor school district in South Africa. Another by Collins et al. (2011) described a 32-condition factorial experiment conducted in primary health care settings in the United States midwest. In some situations conducting a large factorial experiment may not be much more resource-intensive than conducting an RCT, for example, when the intervention is largely Internet-delivered.

When it is scientifically important to examine multiple intervention components, there is a design alternative that enables the investigator to take advantage of the factorial experiment's economical use of subjects while reducing the number of experimental conditions that must be implemented. Fractional factorial designs (Wu and Hamada 2009; Collins et al. 2009; Dziak et al. 2012) are a special type of factorial design with a long history in engineering and related fields, but to date they have rarely been used in the behavioral sciences. In fractional factorial designs, which are associated with exactly the same sample size requirements as complete factorial designs, a carefully selected fraction (hence the name) of experimental conditions is included. A fractional factorial design was used by Strecher et al. (2008) to examine the performance of five components of an online smoking cessation intervention. The design reduced the number of experimental conditions by half, from 32 to 16. A fractional factorial design was used by Collins et al. (2011) to examine six components of a clinic-based smoking cessation intervention. The design cut the number of experimental conditions by half, from 64 to 32. Depending on the number of factors in the experiment, there may be fractional factorial designs available that cut the number of experimental conditions by 75 % or more.

Fractional factorial designs are economical, but there is a tradeoff for this economy. Whenever experimental conditions are removed from a complete factorial design, the number of effects that can be estimated is reduced. As a result, effects that could be individually estimated in a complete factorial design become combined in a fractional factorial design. These effects can no longer be disentangled; they are estimated as a "bundle" (Collins et al. 2009; Chakraborty et al. 2009; Wu and Hamada 2009). In other words, they are confounded (or, to use the term that is more common in engineering, aliased). For every fractional factorial design it is known in advance which effects will be bundled together; in fact, the investigator can select a design strategically so as to bundle the effects of primary scientific interest with effects that are expected to be very small. Typically this means selecting a design in which main effects and two-way interactions are bundled with higher-order interactions. For example, in the fractional factorial design used by Collins et al. (2011), each main effect estimate is actually an estimate of that main effect plus one five-way or six-way interaction, and each estimate of a two-way interaction is actually an estimate of that interaction plus one four-way interaction. Because Collins and colleagues had no reason to expect that the three-way and higher-order interactions would be large, the estimates of main effects and two-way interactions were expected to be approximately equivalent to those that would have been obtained in a much more expensive complete factorial experiment.

Fractional factorial designs with different numbers of factors, different numbers of conditions, and different patterns of bundling of effects have been tabulated and are available in books and on the Internet. One convenient method of selecting a fractional factorial design is to use a computer program such as PROC FACTEX in SAS. A brief tutorial on selecting a fractional factorial design, aimed at behavioral scientists, appears in Collins et al. (2009).

When Subjects are Sampled in Clusters

Sampling subjects in clusters is a reality of much behavioral science. Because the family is a cluster of people, cluster sampling is nearly always used to some extent in family research. In much family research, the clustering of individuals into families is itself of research interest. Thus, each family member as well as relationships within the family may play a unique role in the research, and the extent to which family members are alike or unalike is scientifically interesting.

However, some types of clustering are not of scientific interest and therefore are primarily a nuisance, as when families must be sampled from clusters such as schools, hospitals, health care clinics, and neighborhoods. In the hypothetical example reviewed in this chapter, families might have been sampled from several drug abuse treatment centers. Clustering is important because units within a cluster tend to be more alike than two randomly selected units. That is, families who live within the same community tend to be more alike than two randomly selected families will be. The degree of this similarity is expressed in the intra-class correlation.

When clusters are present, randomization to experimental conditions can take place in two ways (Dziak et al. 2012). In within-cluster randomization, the families within a cluster can be randomized to different experimental conditions. This approach is taken when there is little interaction between families within a cluster, and therefore there is little risk of contamination between experimental conditions. The experiment described in Collins et al. (2011) sampled subjects from health care clinics, and used within-cluster randomization because interaction among clinic patients was minimal. In between-cluster randomization, entire units are randomized to experimental conditions. This approach, taken when there is a risk of contamination, is used frequently in educational research and other research, conducted in educational settings, such as drug abuse prevention (Murray 1998; Raudenbush and Bryk 2002). In one study, schools were assigned to experimental conditions for two reasons (Caldwell et al. 2012). One reason was that two of the components were aimed at teachers, and it was believed that teachers within a school would be likely to share information. The other reason was that one of the components, school environment, was aimed at the school as a unit.

The implications of cluster structure and the choice of within-cluster or between-cluster randomization on factorial experiments are discussed by Dziak et al. (2012). One area in which cluster structure can have an effect is statistical power. When between-cluster randomization is used, the presence of a non-zero intra-class correlation results in what is known as the design effect (Murray 1998). Essentially, this means that as compared to a situation in which individual subjects are randomized, more subjects are required to maintain the same level of statistical power. All else being equal, a larger intra-class correlation and larger clusters (i.e., the same N divided into fewer clusters) produce a larger design effect. It has been shown that the design effect has the same effect on factorial experiments as it does in the standard RCT, and that standard approaches to estimating power are accurate when between-cluster randomization is to be used in a factorial experiment (Dziak et al. 2012). They also showed that the design effect is an issue primarily with between-cluster randomization. When within-cluster randomization is structured so that each unit within a cluster has the same experimental condition assignment probabilities, the design effect is typically negligible.

In addition to its effect on statistical power, cluster structure can impact choice of experimental design. Suppose a scientist who works in the area of school-based drug abuse prevention wishes to conduct a factorial experiment to examine the performance of five intervention components. This experiment will involve $2^5 = 32$ experimental conditions. The investigator has access to enough students to achieve an acceptable level of power, even after taking the design effect into account. However, the students are clustered into 30 schools, and it is determined that between-cluster randomization is necessary. Thus, the investigator has a sufficiently large sample of individual subjects, but does not have enough clusters to implement 32 experimental conditions.

The investigator must somehow reduce the number of experimental conditions. One approach would be to abandon the idea of conducting a factorial experiment and choose a different approach, such as individual experiments. This idea is unappealing for two reasons. First, it would rule out the possibility of examining interaction effects. Second, with a different approach the investigator would have to recruit many more subjects, and therefore many more schools. Another approach would be to reduce the number of intervention components being examined, which would reduce the scientific yield of the study and would probably require the same number of subjects as the larger experiment. Simply reducing the number of intervention components does not necessarily reduce the sample size requirements of a factorial experiment, and, conversely, increasing the number of components does not necessarily increase sample size requirements (See Collins et al. 2009). A third approach would be to retain the same number of intervention components for examination and select a fractional factorial design. This would reduce the number of experimental conditions by at least half and thereby make the experiment feasible with 30 schools. As mentioned above, bundling of effects must be considered carefully when evaluating a fractional factorial design. The use of fractional factorial designs with between-cluster randomization is discussed by Dziak et al. (2012).

Interactions

Two variables are said to interact when the effect of one is different depending on the level of the other. An investigator who raises the issue of interactions usually has one of two different kinds of interactions in mind. The first kind is an interaction between an intervention or intervention component and family characteristics, such as whether or not the family is headed by a single parent, below the poverty line, or of a particular ethnicity. These interactions are of interest because they suggest that an adaptive intervention (Collins et al. 2004; Lei et al. 2012) could be developed that would be tailored to improve performance in certain subgroups. For example, suppose there was evidence of an interaction between MDFT and family ethnicity, suggesting that this intervention component was successful with all families except Hispanics. The intervention package would have a larger overall effect if Hispanic families were provided with a component that better met their needs. The new component could be a specialized version of MDFT or an entirely different component.

The second kind of interaction involves two or more intervention components. In MOST, decisions about which components to retain for potential inclusion in the intervention and which to reject depend largely on main effects. Interactions are examined primarily to determine whether they indicate that these decisions should be reconsidered (Collins et al., in preparation). For example, if the main effect of Academic Skill Building (ASB) disappears when MDFT is included, this suggests that only one of the two components should be included in the intervention.

It may be helpful to clarify an issue about interpretation of main effects when interactions are present. The interpretation of any effects in ANOVA depends on what approach is used for coding the effects. In the behavioral sciences typically either dummy (0,1) or effect (− 1,1) coding is used. Dummy coding produces effect estimates that technically are not main effects and interactions (at least according to the widely accepted textbook definitions, such as those found in Montgomery, 2009; for a detailed explanation, see Kugler et al. 2012). Moreover, these effects are often substantially correlated. Effect coding, which is recommended for component screening experiments (Collins et al. 2009), produces estimates of main effects and interactions that do correspond to the textbook definitions. Moreover, these estimates are orthogonal when the sample size is equal across experimental conditions (and nearly orthogonal otherwise in most cases), making it much more reasonable to consider each effect on its own merits.

Open Areas and Future Directions

One open area is how to make decisions about which intervention components to include in an intervention when some components may have different effects on different outcome variables. In family research different outcome variables may be relevant for different family members. In the drug abuse treatment example, substance use might be the primary outcome for the adolescent, and parental effectiveness might be the primary outcome for the parent. In addition, there might be outcomes that pertain to dyadic relationships within the family, such as effectiveness of parental communication, or outcomes that pertain to the family as a unit, such as amount of family conflict. Decision making can be complicated under these conditions, particularly when experimental results are inconsistent across outcomes, and guidelines are needed. Another open area is how to use the MOST framework to develop the most cost-effective intervention. Additional research is also needed on experimental design, to increase the options available for intervention scientists.

It is interesting to consider what the future direction of family-based interventions might look like if the MOST framework were widely adopted. First, the pace with which the field accumulates a coherent base of scientific knowledge about which intervention approaches work, and which do not, might be accelerated, because every study would shed light on which components were effective. Second, scientists might begin engineering family interventions to meet specific and clearly stated optimization criteria, such as the most effective intervention that can be delivered under some particular dollar limit. Third, once this practice becomes common, the bar could be raised systematically and incrementally over time, so that one objective of new interventions would be to improve on the old by demonstrating that they were more effective, more efficient, or less costly. In this way, behavioral interventions could steadily make ever-increasing progress in improving the lives of families.

Acknowledgments This work was supported by Award Number P50DA010075-15 from the National Institute on Drug Abuse. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institute on Drug Abuse or the National Institutes of Health. The author thanks Amanda Applegate for editorial assistance.

References

Caldwell, L. L., Smith, E. A., Collins, L. M., Graham, J. W., Lai, M., Wegner, L., Jacobs, J. (2012). Translational research in South Africa: Evaluating implementation quality using a factorial design. *Child and Youth Care Forum, 41,* 119–136.

Chakraborty, B., Collins, L. M., Strecher, V. J., & Murphy, S. A. (2009). Developing multicomponent interventions using fractional factorial designs. *Statistics in Medicine, 28,* 2687–2708.

Clapham, C., & Nicholson, J. (2009). Optimization. *The concise Oxford dictionary of mathematics*. www.oxfordreference.com/views/ENTRY.html?subview=Main&entry=t82.e2047. Accessed 3 April 2011.

Collins, L. M., Murphy, S. A., & Bierman, K. (2004). A conceptual framework for adaptive preventive interventions. *Prevention Science, 3,* 185–196.

Collins, L. M., Murphy, S. A., Nair, V., & Strecher, V. (2005). A strategy for optimizing and evaluating behavioral interventions. *Annals of Behavioral Medicine, 30,* 65–73.

Collins, L. M., Murphy, S. A., & Strecher, V. (2007). The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): New methods for more potent e-health interventions. *American Journal of Preventive Medicine, 32,* S112–S118.

Collins, L. M., Dziak, J. R., & Li, R. (2009). Design of experiments with multiple independent variables: A resource management perspective on complete and reduced factorial designs. *Psychological Methods, 14,* 202–224.

Collins, L. M., Baker, T. B., Mermelstein, R. J., Piper, M. E., Jorenby, D. E., Smith, S. S., & Fiore, M. C. (2011). The multiphase optimization strategy for engineering effective tobacco use interventions. *Annals of Behavioral Medicine, 41,* 208–226.

Dziak, J. D., Nahum-Shani, I., & Collins, L. M. (2012). Multilevel factorial experiments for developing behavioral interventions: Power, sample size, and resource considerations. *Psychological Methods, 17,* 153–175.

Fisher, R. A. (1926). The arrangement of field experiments. *Journal of the Ministry of Agriculture of Great Britain, 33,* 503–513.

Hagan, M. J., Tein, J. Y., Sandler, I. N., Wolchik, S. A., Ayers, T. S., & Luecken, L. J. (2012). Strengthening effective parenting practices over the long term: Effects of a preventive intervention for parentally bereaved families. *Journal of Clinical Child and Adolescent Psychology, 41,* 177–188.

Kirk, R. E. (2012). *Experimental design: Procedures for the behavioral sciences* (4th ed.). Los Angeles: Sage.

Kugler, K. C., Trail, J. B., Dziak, J. J., & Collins, L. M. (2012). *Effect coding versus dummy coding in analysis of data from factorial experiments (Technical report No. 12–120).* University Park: The Methodology Center, Penn State.

Lei, H., Nahum-Shani, I., Lynch, K., Oslin, D., & Murphy, S. A. (2012). A "SMART" design for building individualized treatment sequences. *Annual Review of Clinical Psychology, 8,* 21–48.

Liddle, H. A., Dakof, G. A., Turner, R. M., Henderson, C. E., & Greenbaum, P. E. (2008). Treating adolescent drug abuse: A randomized trial comparing multidimensional family therapy and cognitive behavior therapy. *Addiction, 103,* 1660–1670.

McClure, J. B., Derry, H., Riggs, K. R., Westbrook, E. W., St John, J., Shortreed, S. M., An, L. (2012). Questions about quitting (Q2): Design and methods of a multiphase optimization strategy (MOST) randomized screening experiment for an online, motivational smoking cessation intervention. *Contemporary Clinical Trials, 33,* 1094–1102.

Montgomery, D. C. (2009). *Design and analysis of experiments* (7th ed.). Hoboken: Wiley.

Murray, D. M. (1998). *Design and analysis of group-randomized trials*. New York: Oxford University Press.

Prado, G., & Pantin, H. (2011). Reducing substance use and HIV health disparities among Hispanic youth in the U.S.A.: The Familias Unidas program of research. *Psychosocial Interventions, 20,* 63–73.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage.

Sampl, S., & Kadden, R., (2001). *Motivational enhancement therapy and cognitive behavioral therapy for adolescent cannabis users: 5 sessions* (Vol. 1). Rockville: Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration.

Sandler, I. N., Ma, Y., Tein, J. Y., Ayers, T. S., Wolchik, S., Kennedy, C., & Millsap, R. (2010). Long-term effects of the family bereavement program on multiple indicators of grief in parentally bereaved children and adolescents. *Journal of Consulting and Clinical Psychology, 78,* 131–143.

Shaw, D. S., Dishion, T. J., Supplee, L., Gardner, F., & Arnds, K. (2006). Randomized trial of a family-centered approach to the prevention of early conduct problems: 2-year effects of the family check-up in early childhood. *Journal of Consulting and Clinical Psychology, 74,* 1–9.

Stanger, C., Budney, A. J., Kamon, J. L., & Thostensen, J. (2009). A randomized trial of contingency management for adolescent marijuana abuse and dependence. *Drug and Alcohol Dependence, 105,* 240–247.

Strecher, V. J., McClure, J. B., Alexander, G. W., Chakraborty, B., Nair, V. N., Konkel, J. M., Pomerleau, O. F. (2008). Web-based smoking cessation programs: Results of a randomized trial. *American Journal of Preventive Medicine, 34,* 373–381.

Wu, C. F. J., & Hamada, M. S. (2009). *Experiments: Planning, analysis, and optimization*. New York: Wiley.

Chapter 15
Methods in Multi-site Trials of Family-based Interventions

Greg J. Duncan

Methods in Multi-site Trials of Family-based Interventions

The chapters in Sect. IV provide insights into quite a number of methodological issues arising in evaluation research. Most are not unique to family interventions but instead apply more generally to interventions directed either at individuals or groups of individuals. I concentrate on a subset of these issues. The first is the multiple testing problem raised by Moore and Wood (Chap. 13). Multiple testing is endemic to empirical research in the social and behavioral sciences, and the biases it can create are neither emphasized in graduate student training nor accounted for in most of the empirical studies that fill our journals. Here, I point to the advantages of *replication* across experimental sites or across experiments themselves as a very useful antidote for multiple testing bias.

Second, family-based interventions typically target a number of individuals (e.g., both parent and children). In addition to the usual problem of multiple testing bias arising from the many impact estimates that might be gathered about these individuals, there is the added problem of how to weight impacts across different kinds of family members to arrive at some overall judgment of whether a given intervention was successful or beneficial.

My third set of comments arise from the fact that Building Strong Families and many other interventions are conducted in a number of sites, sometimes with identical but often with merely similar "treatments." In addition to providing evidence on the replicability of key findings across sites, these designs can provide powerful means for estimating the magnitude of some of the nonexperimental linkages in hypothesized theories of change. The method for doing this—instrumental variables—has been around for decades, but multi-site experimental studies provide an unusually powerful means of implementing it.

G. J. Duncan (✉)
School of Education, University of California, Irvine, CA, USA
e-mail: gduncan@uci.edu

S. M. McHale et al. (eds.), *Emerging Methods in Family Research,*
National Symposium on Family Issues 4, DOI 10.1007/978-3-319-01562-0_15,

Finally, I point to an underemphasized problem in these chapters—the fact that a given treatment may have very different kinds of effects for different types of families or individuals. While we are developing sophisticated methods for quantifying the scope of this so-called treatment effect heterogeneity, we desperately need theory to help guide us in explaining family circumstances that might lead some families to profit much more than others from a given intervention. Throughout my discussion I draw examples from a set of family-based welfare-to-work experiments conducted by Manpower Demonstration Research Corporation (MDRC) in the 1990s.

Multiple Testing Problems

Multiple testing problems are somewhat easier to handle in program evaluations than in run-of-the-mill empirical analyses of secondary data. In the latter case, analyses often involve a broad search among available measures that is often guided by preliminary results from dozens of regressions of possible outcomes on possible mediators and background measures. Account is rarely taken of the fact that the final hypothesis tests are often suggested as much by the data themselves than by prior theorizing, which leads to rejecting too many null hypotheses.

Program evaluations sometime avoid multiple testing bias. Indeed, many of the original "evaluations" that led the development of statistical tools for analyzing experiments focused on agricultural manipulations of such environmental factors as rainfall and fertilizer (e.g., Bartlett 1937). Manipulations were defined simply and precisely, outcomes were few in number (typically crop yields) and the data themselves were gathered over the course of a growing season. In most cases, data ransacking was not possible since only one kind of statistical analysis could be performed.

Planning modern social interventions and their evaluations is a complicated business, as Moore and Wood (Chap. 13) point out. It typically involves constructing logic models that specify the outcomes targeted by the intervention, the mediational pathways by which experimental impacts might occur, and possible baseline moderators. In theory it is possible to specify in advance a small set of outcomes, mediational paths and baseline moderation, and stick to testing only those. Occasionally, that is what happens. An example is the Infant Health and Development program where its statistician, Helena Kraemer, insisted the study's major publications contain only tests of a relatively small number of pre-specified hypotheses. More commonly, evaluators' and sponsors' spirits may be willing but their flesh is weak. Ideas developed while watching the intervention unfold and during preliminary stages of data analysis can lead to almost as much data mining as in the case of secondary data analysis. Hence, the need for the discipline of adjustments for multiple testing.

The sensible approach advocated by Moore and Wood (Chap. 13) combines some discipline in pre-specifying the set of treatment/control contrasts to be included in the analysis with aggregating, where possible, multiple indicators of latent concepts into indexes, and then employing existing statistical corrections for lingering multiple-testing issues. When they applied the latter to data from the Oklahoma City site of the Building Strong Families (BSF) interventions, they found that a number of the impacts dropped below conventional levels of statistical significance.

Replication as an Alternative Approach to the Multiple Testing Problem

Other approaches to the multiple testing problem might be considered, most notably *replication*; Do patterns generated from one possibly data-mined set of results hold up in a fresh set of data? In the context of evaluations, this may mean using data from some sites to test models generated from other sites. In the case of secondary data analysis, it may involve multiple data sets.

Replication is a key component of the scientific method and a staple in experimental psychology, clinical trials, and many of the natural sciences. Although most replications attempt to reproduce experimental conditions, treatments, and measurement as precisely as possible, there is also great value in testing for replicability under a wider range of designs and contexts. In her classic empirical study of the childhood antecedents of adult antisocial behavior, Lee Robins (1978) emphasized the value of replication across contexts. She wrote: “In the long run, the best evidence for the truth of any observation lies in its replicability across studies. The more the populations studied differ, the wider the historical eras they span; the more the details of the methods vary, the more convincing becomes that replication” (Robins 1978, p. 611). In the case of the Moore and Wood (Chap. 13) study, different versions of the BSF interventions were implemented in a total of eight sites and all shared the same basic elements (group sessions on relationship skills, support from family coordinators, and assessment and referrals to existing community services). The more results hold up across sites, the greater the confidence we can have in them.

My mentor at the University of Michigan, James Morgan, attempted to promote replication among researchers using data from the Panel Study of Income Dynamics. Buried in the PSID’s list of hundreds of variables is one called the “Split-sample filter variable,” which divides the PSID sample into four independent subsamples. Given the complex nature of the sample design, the allocation of cases to the independent quarter samples had to be done by samplers rather than a random-number generator, since whole sampling units had to be assigned to one or another of the quarter samples.

The resulting measure enables a researcher to engage in unbridled data ransacking with a random one-quarter, half, or three-quarters of the sample, and then preserve the remainder for a single, clean estimation of parameters and standard errors of the final model. I know of only one study (Hymans and Shapiro 1974) that followed this protocol precisely, but the idea is a good one. Perhaps a more effective enforcement mechanism (e.g., preserving the “test” sample in a location that researchers can only access when they are ready to use it) is needed.

Evaluation research provides more examples of replication success stories. One comes from welfare-to-work studies conducted in the 1990s (Morris et al. 2001; Duncan et al. 2011). The authors drew data from seven random-assignment studies that collectively evaluated 10 welfare and antipoverty programs in 11 sites, producing a total of 16 program/site combinations: Connecticut's Jobs First; Florida's Family Transition Program; Minnesota Family Investment Program (MFIP tested the effects of two programs—Full MFIP in urban and rural counties and MFIP Incentives Only in urban counties only); Canadian Self-Sufficiency Project (SSP tested the effects of two programs—SSP in the provinces of New Brunswick and British Columbia and SSP-Plus in New Brunswick); the Los Angeles Jobs First GAIN; New Hope; and the National Evaluation of Welfare to Work Strategies (NEWWS tested the effects of two programs—LFA- labor force attachment and HCD-human capital development, in each of three sites: Atlanta, GA, Grand Rapids, MI, and Riverside, CA.)

Across these studies, various packages of welfare and antipoverty policies were tested. All programs were aimed at increasing the self-sufficiency of low-income parents. The programs can be characterized by four program components—earnings supplements, work-first, education/training first, and time limits; most mixed and matched these strategies to implement a "packaged" approach. Earnings supplement programs were designed to make work pay by providing cash supplements outside the welfare system or allowing parents to keep part of their welfare grant as their earnings increased. The remainder of programs were intended to boost work through the use of services, sanctions, and time limits (Gennetian et al. 2002).

From these program evaluations, Duncan et al. (2011) drew data consisting of 18,677 child observations taken from 10,238 children living in 9,113 primarily single-parent families. In the analysis, children's ages ranged from 2 to 5 at the time of random assignment. Impacts on economic and child achievement measures are presented in Table 15.1. The top panel pools studies by (overlapping) types of program; the bottom panel shows study-specific results. In all models, baseline child, parental, and family characteristics are included as control variables.

Looking first at the top panel of Table 15.1, it can be seen that all types of programs boosted post-random assignment employment significantly, with proportions of quarters employed increasing from 4.4 to 8.6 percentage points. Not shown are analogous increases in labor market earnings. In contrast, impacts on family income differed significantly across program types. Earnings supplement programs produced the largest impacts on family income, amounting to nearly $1,500 per year. So-called "work first" programs that relied on mandates rather than earnings incentives had much smaller impacts on family income ($466 per year), in large part because welfare benefits tended to fall almost as much as earnings rose. Programs that prioritized education or job skills and included time limits had similar modest impacts on total family income.

Program-specific details provided in the bottom panel of Table 15.1 show a wider range of impacts, but similar patterns. Program impacts on family income ranged from − $1,000 to over + $2,200, but are statistically significant (and positive) only for the earnings supplement programs. Thus, the earnings supplement approach to increasing family income replicated across the programs built on that strategy.

The employment and income impacts are a prelude to analyses conducted by Morris et al. (2001) and Duncan et al. (2011) that used data from these experiments to test hypotheses about impact of parental employment and family income on children's

Table 15.1 Impacts of Welfare-to-Work programs on parents' economic outcomes and child achievement, by program type and study/site (Duncan et al. (2011), Table 3)

	Proportion of quarters employed		Average annual income (in $1,000s)		Child achievement	
Program Type (not mutually exclusive)						
Earnings Supplements *(n = 8,941)*	0.086	(0.010)***	1.468	(0.158)***	0.076	(0.017)***
Work-First *(n = 9,957)*	0.079	(0.010)***	0.446	(0.156)***	0.040	(0.025)
Education/Training First *(n = 6,017)*	0.044	(0.014)***	0.308	(0.196)	0.027	(0.037)
Time limits *(n = 2,631)*	0.061	(0.016)***	0.505	(0.251)**	0.064	(0.034)
Study/Site						
Self-Sufficiency Project						
British Columbia *(n = 2,131)*	0.076	(0.023)***	1.609	(0.327)***	0.112	(0.050)**
New Brunswick *(n = 1,844)*	0.135	(0.023)***	1.916	(0.274)***	0.125	(0.054)**
SSP Plus-New Brunswick *(n = 471)*	0.116	(0.047)**	2.244	(0.455)***	0.061	(0.113)
Connecticut's Jobs First (n = 1,521)	0.053	(0.021)**	0.813	(0.362)**	0.032	(0.054)
Minnesota Family Investment Program						
Urban Full Program *(n = 1,101)*	0.114	(0.022)***	1.302	(0.409)***	0.045	(0.056)
Urban Incentives Only *(n = 988)*	0.050	(0.024)**	0.914	(0.413)**	−0.016	(0.062)
Rural Full Program *(n = 378)*	0.086	(0.041)**	1.630	(0.585)***	−0.013	(0.093)
New Hope (n = 1,049)	0.065	(0.023)***	1.630	(0.541)***	0.034	(0.066)

Table 15.1 (continued)

	Proportion of quarters employed		Average annual income (in $1,000s)		Child achievement	
NEWWS Programs						
Atlanta LFA *($n = 2{,}324$)*	0.046	(0.023)**	0.317	(0.286)	0.074	(0.059)
Atlanta HCD *($n = 2{,}607$)*	0.020	(0.021)	0.398	(0.262)	0.101	(0.056)*
Grand Rapids LFA *($n = 1{,}458$)*	0.104	(0.027)***	0.072	(0.452)	0.030	(0.070)
Grand Rapids HCD *($n = 1{,}388$)*	0.026	(0.026)	0.004	(0.462)	−0.050	(0.078)
Riverside LFA *($n = 1{,}896$)*	0.121	(0.028)***	0.095	(0.497)	0.007	(0.070)
Riverside HCD *($n = 2{,}022$)*	0.099	(0.024)***	0.557	(0.354)	−0.036	(0.066)
Los Angeles Jobs-First GAIN ($n = 169$)	−0.004	(0.062)	−1.047	(0.884)	0.000	(0.157)
Florida Transition Program ($n = 1{,}110$)	0.069	(0.022)***	0.137	(0.330)	0.112	(0.063)*

Standard errors in parentheses

Means and standard deviations of each variable are as follows: average hours employed per quarter (mean = 145.3, sd = 167.3); average annual income (mean = 11.0, sd = 5.2); child achievement (mean = 0.030, sd = 0.993). Economic variables are measured over study follow-up, child achievement is measured at time of follow-up. Separate regression equations are conducted for each study impact. The regressions also include the following covariates measured at baseline: child's age, earnings in the prior year, earnings in the prior year squared, amount of time mother was on welfare, employed in prior year, mother had no high school degree or equivalent, mother's marital status, number of children in the family, age of youngest child, mother's race/ethnicity, and whether parents' age was less than 18 at the time of child's birth; also included were the following additional covariates: elapsed time between study entry and follow-up, type of achievement report (e.g. parent or test/teacher, when applicable)

*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$ (two-tailed)

achievement. The rhetoric surrounding the welfare reform debate of the 1990s was filled with claims about the importance of role modeling provided by working parents to their children (Chase-Lansdale and Duncan 2001; DeParle 2005). Others claimed that work alone would not be sufficient to improve children's prospects—higher family income, they argued, was the active ingredient for improved child well-being. Some even claimed that increasing maternal work would hurt children by relegating them to substandard child care settings.

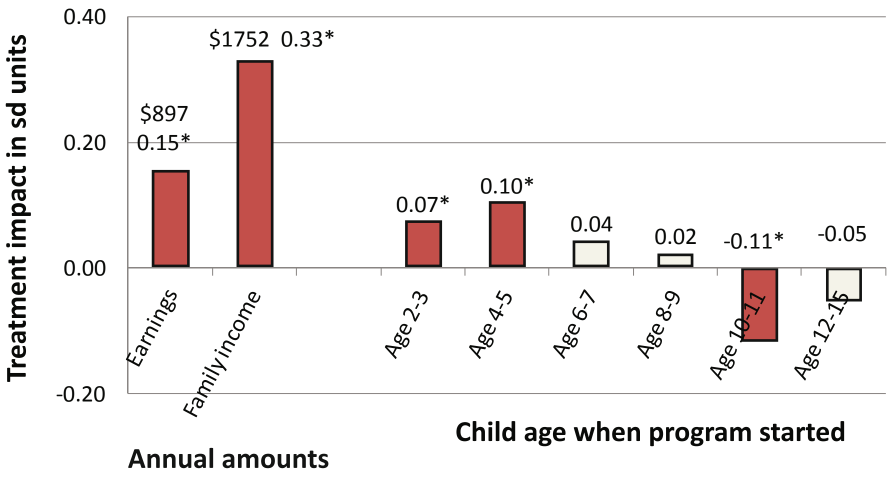
If work alone is what matters for child well-being, then we would expect to see that the experiments that boosted work would also tend to boost children's achievement. Standard-deviation impacts on child achievement are shown in the final set of columns in Table 15.1. A look at the site-by-site results shows that the New Brunswick SSP site indeed produced the largest employment *and* child achievement impacts. However, the site that produced the second largest employment impact (Riverside's labor force attachment treatment stream) produced no improvement in child test scores.

Rather than a site-by-site inspection of these patterns, it is most helpful to aggregate the individual studies by program type (top panel). For this analysis, microdata from the various experiments were harmonized and pooled into a single data set. Results show much greater variation across program type in child achievement impacts than employment impacts. In contrast, child achievement impacts match family income impacts much more closely. Taken as a whole, the earnings supplement programs are leaders in both income gains and in child achievement gains. Thus, the patterns of replication across the experiments shows that family income, not parental employment, appears to be the active ingredient for producing improvements in young children's achievement.

Aggregating Impacts Across Family Members

Virtually all interventions, whether targeting individuals or families, are intended to influence a variety of outcomes. For example, center-based child care might be designed to improve both the cognitive and socioemotional development of children. It is common for interventions to produce larger positive impacts on some outcomes than others. Less commonly, significant impacts may be beneficial for some outcomes but detrimental for others. An example is the case of center-based care, where modest beneficial impacts are often found for cognitive outcomes, while negative impacts are sometimes found for outcomes such as anti-social behavior (Vandell et al. 2010).

How can one judge the *overall* effectiveness of interventions targeting multiple outcomes? The Collins chapter (Chap. 14) identifies this as a fruitful area for future development. Moore and Wood (Chap. 13) cast this as a multiple testing problem—the non-conservative bias in classical tests of statistical significance when a single treatment/control contrast is used to generate significant tests for a host of outcomes. One approach is to combine outcomes into a single index, often by standardizing each component and adding together the standardized scores. This effectively gives



Note: * p<.05; Source: Morris et al. (2005).

Fig. 15.1 Impacts of Earnings Supplement Programs on Mother's Earnings, Family Income and, by age of child, School Achievement

equal weighting to each of the components, and will produce a positive overall impact estimate if the sum of positive impacts exceeds the sum of negative ones.

Family-level interventions can add an additional and perplexing layer of complexity to the index approach. Even if we can calculate a single impact index for each individual in the family, the question is how to weight impacts across family members? Should adults count more than children due to adults' larger contributions to family resources, or should those priorities be reversed because of the longer periods of time over which children, as opposed to adults, may reap their benefits?

An example of this dilemma can be seen in the impacts from four of the random-assignment earnings supplement programs described above (Connecticut's Jobs First; Minnesota Family Investment Program; Canadian Self-Sufficiency Project, and New Hope). The earlier discussion of results from these data (Duncan et al. 2011) focused on children between the ages of 2 and 5 at baseline. In an expanded examination, children between the ages of 2 and 15 at baseline were included (Morris et al. 2005). On average, the annual earnings of program participants exceeded earnings of control-group members by about $900 (one-sixth of a standard deviation). Impacts on family income were similar to those shown in Table 15.1—$1,750 (one-third of a standard deviation).

By promoting maternal work and family income, it was hoped that children would profit from these programs. But in this case, treatment impacts on children's school achievement varied markedly by age (Fig. 15.1). Treatment-group children between the ages of two and five when the programs began, most of whom would be making the transition into elementary school during or shortly after the programs were in operation, scored significantly higher on achievement tests than their control group counterparts, with the achievement differences amounting to 0.07 to 0.10 sd. The achievement of children age 6–9 did not appear to be affected by the programs. Surprisingly, the achievement of children who were age 10 and 11 when the programs began seemed to be hurt by the programs' efforts to increase parental employment and family income.

Whatever the causes (about which we speculate below), the collection of both positive and negative impacts makes it difficult to render an overall judgment on whether these programs were worthwhile. Despite a number of significant beneficial impacts, the fact that children making the transition into adolescence appeared to be hurt by the program forces one to develop some kind of weighting scheme across individuals in a family. Does equal weighting across individuals risk the possibility that the formidable costs associated with the increased chances of producing high-risk teens outweigh achievement gains for younger children and employment gains for mothers? They may well. The dollar value to taxpayers of saving a youth from serious crime and drug abuse and from becoming a high school dropout was estimated by Cohen (1998) to amount to millions of dollars.

Benefit-cost analysis, where dollar values of each impact are estimated, summed, and then compared with the intervention's cost (Gramlich 1990), provides a natural way around this problem. Adolescent risks are weighted by their estimated dollar values, while the benefits (e.g., less grade failure, lower chances of needing special education) of achievement gains for younger children and employment gains for mothers are weighted by their magnitude as well. Of course, benefit-cost approaches are not without problems, the most prominent of which is that not all benefits can be monetized.

Using Multiple Experimental Data Sets to Estimate Impacts of Mediators on Outcomes

Most interventions are based on theories of change that account for the processes by which the experimental condition's direct and indirect effects ultimately affect the outcomes of interest. An explicit example of this in the case of the Building Strong Families initiative can be seen in Fig. 13.1 (Chap. 13). In brief, the BSF treatment provides services that are presumed to improve couple relations, then parenting and, ultimately, child well-being.

A discussion of the advantages (and complications) introduced by using a MOST-type design to evaluate the underlying theory of change would be a valuable addition to the Collins chapter (Chap. 14). According to Collins, the first stage of MOST generates a set of impact estimates associated with different treatment arms. But if process model mediators were measured, then the first stage could also generate impact estimates on those mediators. Intervention designers might think differently about MOST results if treatment components affect ultimate outcomes but not the hypothesized pathways by which those outcomes were thought to be influenced by the treatment.

Random assignment experiments provide strong causal evidence on the impacts of the treatments on measures drawn from each of the stages in the theory of change

(Collins, Chap. 14). In the case of BSF, this amounts to estimates of impacts on services received, couple relations and on child outcomes. These patterns of impacts provide indirect evidence on the theory of change. If, for example, a BSF intervention improved child outcomes but failed to affect parenting, then we might doubt that parenting was an active ingredient in conveying the impacts of the interventions on child outcomes.

Most needed in testing theories of change is some way of estimating all of the hypothesized paths in one's causal chain that, in the case of BSF, would include the impacts of couple relations on parenting and the impacts of both couple relations and parenting on child outcomes. But since couple relations and parenting were not subject to experimental manipulation, conventional methods for estimating these links using data from experiments are subject to the same kind of omitted variable bias as are estimates from more conventional survey data. In fact, using data drawn from experiments for this purpose tends to be worse than using more conventional sources of survey data since samples for the latter are often drawn to represent populations of interest whereas samples enrolled in experiments tend to be demographically peculiar.

Experimental data taken from multiple sites provide a possible avenue for securing much stronger estimates of the nonexperimental paths of process models, using the method of instrumental variables (Gennetian et al. 2008). The most complete published example of this is based on data from the Moving to Opportunity residential mobility experiment, which are used to estimate the impact of neighborhood conditions such as the poverty rate on youth outcomes (Kling et al. 2007; Ludwig and Kling 2007). Here we describe Duncan et al.'s (2011) use of data from the welfare-to-work experiments described above to estimate the impact of family income on child achievement.

The basic idea is simple: Assuming the goal is to obtain an unbiased estimate of the effects of income on child achievement and that one can isolate a portion of variation in family income that is *unrelated* to unmeasured confound variables (e.g., parent competence; environmental risks), then using *only* that portion to estimate income effects, the resulting estimates of income effects on child achievement are likely to be free from omitted-variable bias.

In the case of Duncan et al. (2011), random assignment to treatment or control groups in the various welfare experiments is an excellent candidate for an IV variable since families had no say in whether they were assigned to experimental or control groups. More formally, the method uses interactions between treatment group assignments (T) and sites (S) as instrumental variables to isolate experimentally-induced variations in income (Inc) and achievement (Ach) across program models. Using X to denote baseline covariates, the two-stage model is:

$$\mathrm{Inc} = \mathrm{T}^{*}\mathrm{S}\gamma_1 + \mathrm{S}\gamma_2 + \mathrm{X}\beta_1 + \varepsilon_1 \quad (15.1)$$

$$\mathrm{Ach} = \mathrm{PredInc}\lambda_1 + \mathrm{S}\lambda_2 + \mathrm{X}\beta_2 + \varepsilon_2 \quad (15.2)$$

In effect, Eq. (15.1) estimates the family income (Inc) of each child based on experimental assignment and controls by site, while Eq. (15.2) relates the *predicted* level of Inc (PredInc) for a given child taken from (15.1) to that child's observed achievement (Ach). The inclusion of the same control variables in (15.2) as in (15.1) and, importantly, of site fixed effect dummies in both equations, ensures that the only variation in Inc used in the estimation of Ach comes from the lottery-based assignment to treatment and control groups by site. The success of IV models such as (15.1)–(15.2) depends on the strength of first stage prediction of income based on random assignment (Bound et al. 1995). Other assumptions are important as well. See Duncan et al. (2011) for a more complete discussion of these.

The full power of the IV approach comes from pooling data across all studies since that can leverage the variation in impacts on income and child achievement across the studies and sites. If income matters for child achievement, we would expect that the treatment group/site combinations with the biggest positive income deviations should also have the biggest positive achievement deviations. When a trend line is fit through the 28 treatment-site observation points, the slope of the line (0.060) is equal to the IV estimate of the effect of income on child achievement including only site dummies as covariates and using site-treatment interactions as instruments (Kling et al. 2007). The interpretation of the slope is that each $1,000 increase in income is associated with a 0.06 sd increase in child achievement. This may seem like a modest amount, but existing policies such as the Earned Income Tax Credit transfer more than $5,000 per year to some families. A $3,000 income increase translates into a 0.18 sd increase in achievement, an impact comparable to that found in the Tennessee Star experiment with smaller class sizes (Krueger and Whitmore 2001).

In their various IV models, Duncan et al. (2011; Table 5) estimate achievement impacts of income to range from 0.049 sd to 0.062 sd per $1,000, with standard errors ranging from 0.016 to 0.019. This process of using site-specific variation in treatment and control groups can be generalized to other nonexperimental paths in an intervention's theory of change.

Needed: Theories of Treatment Effect Heterogeneity for Families. The Collins chapter does a very nice job of laying out the case for a two-step process of intervention design if program designers are not sure of the comparative effectiveness of several treatment arms is laid out well by Collins (Chap. 14). In the first step, a multi-factorial design identifies the most potent treatment elements, while the second provides a conventional treatment/control RCT test of the most promising treatment bundle. It is good to be reminded of the advantages of a multi-factorial experimental design.

Throughout their analyses, both Collins (Chap. 14) and Moore and Wood (Chap. 13). make the common and very convenient assumption that treatment impacts are invariant across the treated individuals and families. In the "Future Directions" section, Collins recognizes that a given treatment may affect family members differently. If the disparate child impacts shown in Fig. 15.1 are at all common, then it is vitally important to understand what has come to be called treatment effect heterogeneity.

Although policy researchers have long recognized that the impacts of programs differ across individuals, disciplinary barriers prevent many from understanding the sources of this variation. Economists and, increasingly, sociologists have developed methods for quantifying the scope of heterogeneous treatment impacts (Imbens and Angrist 1994; Xie et al. 2011). But most economic theories make few concrete predictions regarding the causes of that heterogeneity or the processes producing the impacts inside of the black policy box. And while sociologists offer a sophisticated conception of the many contexts (e.g., neighborhoods, schools) in which children develop, they rarely link those conceptions to the characteristics of children and families within a given context that lead some to profit much more than others from interventions.

In the case of treatment heterogeneity across individuals, Duncan and Vandell (2012) attempt to explicate ways in which educators, intervention scientists and policymakers can make use of neurologically and developmentally informed principles to increase the likelihood that educational programming and behavioral interventions will achieve their intended goals of supporting positive development in children and youth. Key to successful interventions, they argue, is an understanding of how the cognitive and socio-emotional development of all children and adolescents follows a predictable pattern of stages in which windows for profitable interventions widen or narrow. Moreover, successful interventions may also rest on understanding predictable variation *within* a given stage in family circumstances and in the timing of normative development as children reach important developmental milestones (e.g., the ability to focus attention on learning tasks for young children and puberty for adolescents).

More specifically, Duncan and Vandell (2012) generate hypotheses regarding the likely impacts of these policies across and within children's developmental stages by following past work in focusing on the *congruence* ("fit") between the developmental needs of children and youth and the design and nature of intervention policies (Eccles et al. 1993). Children profit most from interventions that are well matched to their developmental stage, individual characteristics, and family circumstances.

For example, an important principle of human development is that although beneficial changes are possible at any point in life, interventions early in life may be more effective at promoting well-being and competencies compared with interventions undertaken later in life. Emerging evidence from human and animal studies highlights the critical importance of early childhood for brain development and for establishing the structures that will shape future cognitive, social, emotional, and health outcomes (Knudsen et al. 2006). An economic model of development in which preschool cognitive and socio-emotional capacities are key ingredients for human-capital acquisition during the school years is proposed by Cunha et al. (2010). In their model, "skill begets skill"; for example, school-entry capacities can affect the productivity of school-age human-capital investments. If most K-12 schooling is geared toward boosting the skills of children meeting normative developmental milestones, and if many children growing up in disadvantaged circumstances begin school well behind their peers, then the potential payoffs to preschool programs boosting early skills may be large indeed. For these reasons, the "fit" between high-quality center-based child care and the preschool developmental period may generate exceedingly high payoffs for children's long-run well-being.

Sometimes, well-intentioned policies provide ill-fitting environments for development. For example, the hypersensitivity of adolescents to peer approval can produce virulent consequences among groups of deviant youth. Studies of the micro-dynamics of deviant groups show how talk of deviance often elicits approving reactions from peers, which in turn leads to more such talk. The resulting reinforcing process has been called "deviancy training" and can subvert well-intentioned programs that provide services to groups of deviant youth (Dishion et al. 1999; Dodge et al. 2006). The problem appears to be most acute in the case of moderately deviant youth—young adolescents in the early stages of manifesting serious behavior problems who lack the family or other supports that might otherwise keep their behavior in check.

These "fit"-based ideas about treatment heterogeneity based on the circumstances of individuals and their match to the nature of treatments need to be generalized to the family level. What characteristics of families, in combination with the nature of the interventions, are likely to produce the biggest and smallest impacts? For example, in the case of the Building Strong Families intervention, what theories can we use to predict for which families the program is likely to be particularly effective or ineffective?

An individual-based example of theory-based approaches to treatment effect heterogeneity comes from the Morris et al. (2005) data displayed in Fig. 15.1. The same collection of welfare-to-work programs that produced favorable impacts on the school achievement of children aged 2–5 at baseline produced negative impacts on children aged 10–15 at baseline. What accounts for the heterogeneous pattern of treatment impacts across these children?

It has been posited that features of children's normative development account for many of these differences (Morris et al. 2005). In the case of children who were transitioning into elementary school, employment and income supports for parents fit well with the children's full-day, reliable, and structured school environments. However, qualitative research revealed that adolescents with younger siblings were burdened by child care responsibilities when their mothers increased their employment (Gennetian et al. 2004).

A vivid example of the latter comes from a companion qualitative research project concerning an adolescent whose mother needed to start work at six o'clock in the morning (Gennetian et al. 2004). The adolescent was responsible for taking her younger sibling to school in the morning, but doing that meant that the adolescent herself could not make it to her school on time. Her no-excuses high school sent her to detention every day. Indeed, additional analyses of the experimental data showed that negative impacts were concentrated among adolescents with younger siblings. Overall, it appeared that work-focused anti-poverty programs fit the developmental needs and family circumstances of younger children much better than older children.

Another example of treatment effect heterogeneity comes from the New Hope experiment and is mentioned briefly by Weisner (Chap. 10). As background, New Hope provided a set of work supports for full-time workers—parents and nonparents, men and women—that would lift them out of poverty as well as provide subsidies for essential benefits in the form of health insurance and child care subsidies for people who needed them. If a participant was unable to find a job due to lack of work experience or a criminal record, for example, the program provided opportunities for temporary community service jobs (CSJs) that paid the minimum wage but still entitled that person to program benefits. Taken together, New Hope offered a cafeteria of benefits from which participants could choose—a feature that could allow families with diverse needs and circumstances to tailor the program to their own unique situations.

Qualitative interviews suggested important heterogeneity in the labor supply responses of mothers among the experimental families. Some, perhaps one-fifth, appeared to have so many problems (e.g., drug dependence, children with severe behavior problems, abusive relationships) that New Hope's package of economic benefits was unlikely to make much of a difference in their lives. A second group comprising two-fifth of New Hope families, was at the other end of the spectrum: they had no such apparent problems and were able to sustain employment on their own without the need for New Hope's work support. In other words, for this group, control families might be expected to do so well in Milwaukee's job-rich environment that it would be difficult for comparably unconstrained experimental families to do better.

A third group, however, with only one of the problems of the sort that New Hope might be able to address (e.g., difficulties in arranging for childcare, a minor criminal record that experience in a community-service job could overcome) appeared poised to profit from the New Hope package of benefits. Extensive quantitative work on subgroups defined according to the number of potential employment-related problems they faced at the beginning of the program confirmed the wisdom of these qualitatively derived insights. Using data gathered from the baseline interviews, Magnuson (1999) constructed an index of potential employment barriers based on past history of employment, completed schooling, arrests, and the presence of either many or very young children. She then estimated treatment-control differences for subgroups defined according to whether the family faced zero, one, or two or more barriers to employment. The results are presented in Fig. 15.2.

Figure 15.2 confirms what she had suspected based on qualitative evidence. Experimental members who had either no barriers or multiple barriers did not earn a significantly different amount than their control counterparts. Program impacts on the earnings of families with only one barrier, however, were large and statistically significant in both years.

Both of these success stories for understanding treatment effect heterogeneity involve individual behavior—children's achievement and mother's work hours. Can these results generalize to family-based interventions focused on family-based outcomes such as couple or parent-child relations? What theories from family sociology or psychology can generate useful predictions regarding the kinds of families most likely to be helped by family-based interventions? If the circumstances of substantial numbers of families lead them to be unresponsive to interventions, can they be identified and targeted with different, more efficacious interventions? Only by combining strong family-based theory with the trench-warfare of real-world intervention can we hope to design policies that promote positive outcomes for all families.

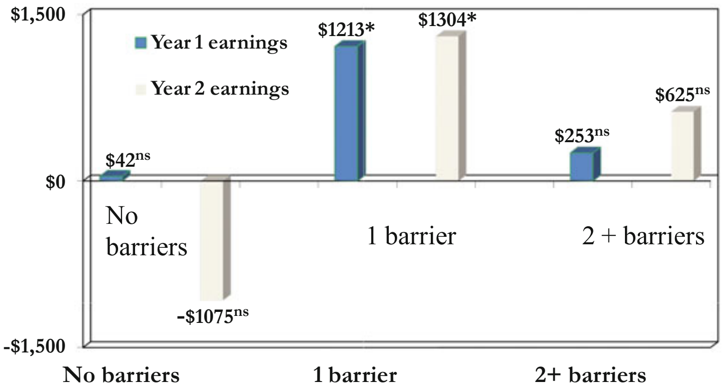


Fig. 15.2 Impacts of New Hope on 1st and 2nd Year Earnings, by Number of Potential Employment Barriers

Conclusion

The chapters in section IV of this volume show how far we have come in developing, refining, and optimizing interventions designed to help either individuals or families. Further progress can be aided by a somewhat broader view of intervention approaches coming out of both academia and policy evaluation firms. Directing interventions at families rather than individuals can complicate things, but in largely surmountable ways.

References

Bartlett, M. (1937). Some examples of statistical methods in research in agriculture and applied biology. *Supplement to the Journal of the Royal Statistical Society, 4*(2), 137–183.

Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association, 90,* 443–50.

Chase-Lansdale, L., & Duncan, G. (Eds.). (2001). *For better and for worse: Welfare reform and the well-being of children and families*. New York: Russell Sage.

Cohen, M. (1998). The monetary value of saving a high-risk youth. *Journal of Quantitative Criminology, 14*(1), 5–33.

Cunha, F., Heckman, J. J., & Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica, 78*(3), 883–931. doi:10.3982/ECTA6551.

Deparle, J. (2005). *American dream: Three women, Ten kids, and a nation's drive to end Welfare*. New York: Penguin Books.

Dishion, T., McCord, J., & Poulin, F. (1999). When interventions harm: Peer groups and problem behavior. *American Psychologist, 54*(9), 755–764.

Dodge, K., Dishion, T., & Lansford, J. (2006). *Deviant peer influences in programs for youth: Problems and solutions*. New York: Guilford Press.

Duncan, G., & Vandell, D. (2012). *Understanding variation in the impacts of human capital interventions on children and youth*. Irvine Network on Individuals in Development. Manuscript in preparation.

Duncan, G., Morris, P., & Rodrigues, C. (2011). Does money really matter? Estimating impacts of family income on young children's achievement with data from random-assignment experiments. *Developmental Psychology, 47*(5), 1263–1279.

Eccles, J., Midgley, C., Wigfield, A., Buchanan, C., Reuman, D., Flanagan, C., & Iver, D. M. (1993). Development during adolescence: The impact of stage-environment fit on young adolescents' experiences in schools and families. *American Psychologist, 48,* 90–101. doi:10.1037/0003-066X.48.2.90.

Gramlich, E. (1990). *Cost-benefit analysis*. New York: McGraw-Hill.

Gennetian, L. A., Huston, A. C., Crosby, D. A., Chang, Y. E., Lowe, E. D., & Weisner, T. (2002). *Making child care choices: How welfare and work policies influence parents' decisions*. New York: MDRC.

Gennetian, L., Duncan, G., Knox, V., Vargas, W., Clark-Kauffman, E., & London, A. (2004). How welfare policies affect adolescents' school outcomes: A synthesis of evidence from experimental studies. *Journal of Research on Adolescence, 14*(4), 339–423. doi:10.1111/j.1532-7795.2004.00080.x.

Gennetian, L., Magnuson, K., & Morris, P. (2008). From statistical associations to causation: What developmentalists can learn from instrumental variables techniques coupled with experimental data. *Developmental Psychology, 44*(2), 381–394.

Hymans, S., & Shapiro, H. (1974). Five thousand American families—Patterns of economic progress. Ann Arbor: Institute for Social Research.

Imbens, G., & Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica, 62*(2), 467–475.

Knudsen, E., Heckman, J., Cameron, J., & Shonkoff, J. (2006). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences*, *103*, 10155–10162. doi:10.1073/pnas.0600888103.

Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica, 75,* 83–119.

Krueger, A., & Whitmore, D. (2001). The effect of attending a mall class in the early grades on college-test taking and middle school test results: Evidence from project STAR. *The Economic Journal, 111,* 1–28.

Ludwig, J., & Kling, J. (2007). Is crime contagious? *Journal of Law and Economics, 50,* 491–518.

Magnuson, K. (1999). Appendix K—The barrier indicator index. In J. Bos, A. Huston, R. Granger, G. J. Duncan, T. Brock, & V. McLoyd (Eds.), *New hope for people with low incomes: Two-year results of a program to reduce poverty and reform welfare* (pp. 346–355). New York: Manpower Demonstration Research Corporation.

Morris, P., Huston, A., Duncan, G., Crosby, D., & Bos, J. (2001). *How welfare and work policies affect children: A synthesis of research*. New York: MDRC.

Morris, P., Duncan, G. J., & Clark-Kauffman, E. (2005). Child well-being in an era of Welfare reform: The sensitivity of transitions in development to policy change. *Developmental Psychology, 41*(6), 919–932.

Robins, L. N. (1978). Sturdy childhood predictors of adult antisocial behavior: Replications from longitudinal studies. *Psychological Medicine, 8,* 611–622.

Vandell, D. L., Belsky, J., Burchinal, M., Steinberg, L., & Vandergrift, N. (2010). Do effects of early child care extend to age 15 years? *Child Development, 81,* 737–756.

Xie, Y., Brand, J., & Jann, B. (2011). *Estimating heterogeneous treatment effects with observational data*. Working paper, Population Studies Center, University of Michigan.

Part V
Conclusions

Chapter 16
Capturing the Complexity of Families Using Innovative Methods

Melissa A. Lippold and Catherine B. McNamee

Studying families poses unique challenges to researchers, including the development and use of innovative methods, such as those outlined in this volume, to capture the complexity of family dynamics. Life course and family systems theories (Elder 1998; Cox and Paley 1997), hold that family members are interconnected; as such, family researchers require methods that capture family interactions and illuminate how the experiences of one family member spillover to affect other individuals in the family. Consistent with ecological models (e.g., Bronfenbrenner 1979), the systems tenet that families are open to external influences means that family processes must be studied as they emerge and develop within a broader ecology. To understand family dynamics, investigators must capture not only the influences of within-family changes, such as developmental changes in family members, but also complex and reciprocal influences between and among family members and between families and their environments. Adding to the challenges facing family researchers, rapid demographic changes have led to increasing variation in family structure, requiring researchers to find new ways to define the family unit (Cherlin 2010).

Theories that pertain to the role of family dynamics in development highlight complex interconnections among family members. Elder's life course theory for example, posits that family members live linked, interdependent lives, such that the behaviors and actions of one family member subsequently affect other members of the family (Elder 1998). Similarly, developmental and ecological perspectives hold that family members form adaptive regulatory systems (Cox and Paley 1997; Minuchin 1985). From a family systems perspective, changes in one family member's behavior may influence the larger family system through crossover effects on other

M. A. Lippold (✉)
Department of Human Development and Family Studies,
The Pennsylvania State University, University Park, PA, USA
e-mail: mal394@psu.edu

C. B. McNamee
Population Research Institute, The Pennsylvania State University,
University Park, PA, USA
e-mail: cbm16@psu.edu

S. M. McHale et al. (eds.), *Emerging Methods in Family Research,*
National Symposium on Family Issues 4, DOI 10.1007/978-3-319-01562-0_16,

family members (Cox and Paley 1997). From a systems perspective, families are best studied in their totality, as the characteristics of the family system are emergent and more than the sum of their component parts. All of these perspectives encourage researchers to develop methods that capture dynamic processes occurring between and among family members and both the internal and external forces that lead to systems-level change.

Bronfenbrenner's ecological model (1979) focuses on influences on individual development, but also provides important insights into studying families. From this perspective, families are embedded within broader ecological contexts, and exchanges involving different levels of the ecosystem serve as "engines" of development and change within the family system. Family systems are likely influenced by factors in the broader ecology. For example, the experiences in the workplace may influence the development of a particular family member, and may subsequently spillover to other family members and change the family system (Repetti et al. 2011). Broader social changes such as changes in norms regarding marriage may influence the choices that a couple makes about their marital status (Amato et al. 1979; Thornton et al. 2007). Bronfenbrenner's theory proposes that researchers use a range of methods to capture interactions between family members and the ecological system and how these interactions may influence the family system.

Demographic analyses also highlight the complexity of families and how the study of families has become increasingly complicated: Recent changes in the structure of families make it increasingly challenging to define the members and roles in a family (Amato et al. 1979; Cherlin 2010). For example, increases in the divorce rate and non-marital births have led to dramatic increases in the number of single parent families. In addition, increases in cohabitation mean that more couples are involved in childrearing outside of marriage. In short, defining a family has become more complicated than simply assessing whether individuals are legally or biologically related. These demographic trends challenge researchers to use new methods to capture increased variation in family structure and what changes in structure mean for family dynamics and systems.

Recent advances in research methodologies described in this volume provide new opportunities to capture the complexity of families that is inherent in systems-oriented perspectives. Novel methods provide opportunities to capture demographic trends and changing family structures and roles. Methodological advances allow scholars to capture dynamic processes and interactions of family members and to broaden the unit of analysis beyond individuals to dyads and whole families. New technologies allow researchers to measure processes occurring at multiple levels of family and ecological systems, including those that occur "under the skin" at the physiological level. Highlighted in this volume are methods that allow researchers to delve deeper into family relationships and to capture more of the complexity of interpersonal dynamics. Also described are innovative methods that can be used to improve the design and evaluation of programs and policies aimed at enhancing family relationships.

This concluding chapter is organized around five themes that emerge across the chapters in this volume: defining families, assessing variation and change in context, overcoming challenges, capturing the big picture, and conducting research with a translational and public health impact. The first section delves into the complexity inherent in family relationships and addresses some of the methods that enable researchers to define "family." We review issues involved in identifying family members and roles and how methods may be applied to capture differing units and levels of analysis (e.g., individual, dyad, whole family system). We also discuss why it may be important to integrate the perspectives of multiple family members. In the second section, we describe methods for capturing variation and change in families as these unfold within a broader ecological context. We highlight ways in which methods provide for distinct conceptualizations of variation and change and influences on change processes. In the third section, we describe some of the challenges scholars face when conducting research on families and some of the innovative methods that have been used to overcome these challenges. In this section we review how selecting the appropriate research design, using multiple and novel methods, and balancing tradeoffs can aid researchers in planning studies that best capture complex and changing family processes. In the fourth section, we describe some of the ways researchers can integrate methods and reach across disciplines to develop a more complete picture of families. In the fifth section, we consider how the design and implementation of research aimed at programs and policies for families may impact public health. We end by discussing the implications of new methods for the field of family studies. We argue that innovative methods are valuable tools in furthering our understanding of family dynamics, provided that researchers remain thoughtful in implementing methods best suited for answering meaningful theoretical questions.

Capturing the Complexity of Families

Families are characterized by interactions and relationships between and among multiple individuals. Many chapters in this volume describe innovative methods that further our ability to characterize families. And, many authors suggest new ways of defining membership in a family and consider new ways to identify the appropriate unit of analysis.

Defining Family Members and Family Roles

Changes in rates of divorce, cohabitation, and non-marital births in the USA have led to increased complexity in the structure of families (Cherlin 2010). Children grow up in families with a range of adults and other children, some of whom are biologically related, some with whom they have legal ties, and some with whom they have less well-defined connections. For example, children may live with both biological parents, a single biological parent, step parents, and cohabiting adults ranging from a parent's romantic partner to grandparents, aunts, and uncles. Children can also live in more than one household as in dual custody arrangements. Children may have sibling and sibling-like relationships with half siblings, cousins, or step siblings; in some families, biological siblings may take on parental roles. Changing family structure has made it increasingly difficult to define the members of a family and to understand the roles of family members.

As Gauthier and Moody (Chap. 5) discuss, traditional definitions of family and traditional research methods may not fully capture new family structures. According to these authors, instead, families may be defined by behaviors that occur between individuals that fulfill particular family roles. These roles may map onto biological relationships, but not necessarily. Gauthier and Moody use social network methods to identify family roles "from the ground up." For example, they use time diary data to define roles based on the types of activities individuals engage in and the nature of the interactions between individuals. Network models hold great potential to help us understand how individuals with a range of relationships—biological, legal, and otherwise—fill certain family roles, and how concepts of family may need to be broadened to account for contemporary diversity in family forms.

The need to distinguish between the structural characteristics of a family and aggregate or relational characteristics is highlighted by Amato (Chap. 11). According to Amato, researchers often measure families based on who is in a family, what are their roles, and what is the nature of their relationships (adoption, biological, marriage). When defining family, researchers often impose their own definitions that may be useful in understanding trends over time. However, Amato posits that researchers may benefit from asking family members who they consider to be their family. Subjective perspectives may result in a more varied and broader definition of family, one that may include relationships such as friends, may span across households, and may be asymmetrical across "members." He argues that using subjective definitions may allow us to map family social networks from a new perspective, enabling the capture of fragmentation among networks, links between networks, and the transmission of social capital. Weisner (Chap. 10) echoes some of these ideas, arguing that qualitative methods may be particularly important in understanding family roles, as researchers can capture subjective experiences and the meanings of social interactions and activities. Such methods may allow researchers the opportunity to discover new ways of conceptualizing family roles and relationships.

The Unit of Analysis

When studying families, researchers are faced with many different potential units of analysis: individuals, dyads, and larger social groups, including whole family units. Families are made up of individuals; however, families also include dyads, such as husband-wife or mother-child. The unit of analysis may extend beyond the individual to capture processes that occur between individuals, such as how members of a dyad respond to and react to one another. Furthermore, families may extend beyond dyads–the entire family unit constitutes a regulatory system, and the whole family has been conceptualized as an important unit of analysis by family systems theorists (Minuchin 1985). As Amato (Chap. 11) and Emery (Chap. 7) point out, many family theories highlight the importance of capturing interdependent processes between and among family members, yet few measures and methods allow researchers to specifically test propositions from family theories.

Several chapters in this volume aim to capture dyadic processes that occur between family members. For example, Teachman (Chap. 1) describes how a researcher might use growth curve models to understand change over time in marital satisfaction. These models can be extended to understand dyadic processes—that is, how members of a dyad, such as a husband and wife, may affect each other's marital satisfaction over time. Teachman found that the pattern of change in one spouse's marital dissatisfaction (i.e., linear change, or slope) was positively correlated with change in the other spouse's dissatisfaction. By allowing for the covariance between the slopes to vary, these analyses can capture the influences that members of a dyad have on each other. State space models have been used by Chow, Messinger, and Mattson (Chap. 3) to illuminate how mothers' and infants' behaviors may influence one another over time during the Still Face Procedure (Adamson and Frick 2003; Tronick et al. 1978). These models can capture how mothers and infants may influence one another's ability to self-regulate. Regulatory functions within a family system and the extent to which spouses return to homeostasis in their moods after external events, such as childbirth are explored by Ram, Shiyko, Lunkenheimer, Doerksen, and Conroy (Chap. 2). Within this framework, the unit of analysis is the couple system and model estimates are dyadic-level estimates. These models allow for family processes to be understood, not only at the individual level, but at the dyadic level as well.

New methods for capturing the family system and changes that may impact family processes are suggested in this volume. By exploring differences in genetic density across different family structures and clusters, Gauthier and Moody (Chap. 5) aim to capture the whole family system. Qualitative methods are described by Weisner (Chap. 10) as the approach to help us understand how whole family systems might change to accommodate one family member. For example, the family system may accommodate a member with a disability when choosing which activities to participate in or by changing family goals. Observational coding methods that capture dyadic and triadic interactions are outlined by Cummings, Bergman, and Kuznicki (Chap. 6) and Metzler, Sanders, and Rusby (Chap. 12). Each of these approaches provides new techniques for studying families as systems, and in describing the approaches the authors point to novel research questions that can be addressed by family researchers.

Capturing Different Family Members' Perspectives

Many studies of families rely on reports from, or measures of, individual family members. However, as Amato (Chap. 11) points out, differences in family members' perspectives of their shared relationships also provide important information. For example, parents may be more likely to view family relationships more positively than their teenagers do (Smetana 1988). Some recent work suggests that differences in parent and child perspectives on parental knowledge of youth activities may be linked to risky behavior (Lippold et al. 2011). These differences may be masked when family processes are indexed with mean scores across family members; measures that tap into differences in family members' perspectives may provide new information (Amato, Chap. 11). Family members often have unique perspectives and experiences, and researchers should not expect agreement between and among reporters (Metzler et al., Chap. 12).

Authors in this volume suggest methods to capture the range of perspectives on family processes. Qualitative data can allow researchers to understand how different family members view and interpret their shared world (Weisner, Chap. 10). For example, Weisner summarizes research that captures both parent and child perspectives on family life when a child has a disability. Observational measures are proposed by Cummings et al., (Chap. 6) to provide important information on family interactions, information that may not be captured using interview or questionnaire methods. This research team uses innovative strategies, such as The MacArthur Story Stem Battery, to collect child narratives on parent child relationships. The children are provided with puppets or dolls, which they use to act out and narrate stories about their attachment relationships and conflict patterns within their family. Taking another tack, Smyth and Heron (Chap. 9) argue that Ecological Momentary Assessment (EMA) data provide opportunities to study different family members' perspectives about the same interactions or experiences. In the domain of applied research, Duncan (Chap. 15) argues that it is important to assess each family member's response to an intervention, because there may be heterogeneity in treatment effects. For example, an intervention may improve outcomes for young children, but have negative effects for older siblings. Measuring the experiences and perspectives of different family members may provide new insights about family processes and lead to a better understanding of why intervention programs work to promote individual and family well-being, and why they sometimes do not.

Capturing Changes in Context

A main focus of developmental research is to understand processes of change over time and the factors associated with such changes. Many chapters in this volume describe repeated data collections of individual and family characteristics and experiences. A key difference among the chapters is the timescale of data collection, with some authors exploring changes that occur at the moment-to-moment level (e.g., Chow et al., Chap. 3), and others exploring changes over larger timescales, such as across years (e.g., Teachman, Chap. 1). Chapters also differ in their focus on (mean) level changes over time versus intra-individual variability. Another key focus was attention to the role of context: Several chapters focus on ways of studying social processes in their natural environments and understanding how contextual influences might moderate change processes. In this second section, we review how authors use novel methods to capture changes over time as they unfold in varying ecological contexts.

Differing Time Scales

The time scale of data collection has implications for the types of research questions an investigator can address (Fuligni, Chap. 4). Some events may unfold rapidly in time whereas others may change more slowly over long intervals of time. A key point by all of the authors of this volume is that investigators should use methodologies matched to the time course of the phenomena under study.

Methods to investigate change processes occurring in relationships and families over short durations of time were discussed in a number chapters (e.g., Chow et al., Chap. 3; Fuligni, Chap. 4; Ram et al., Chap. 2; Smyth and Heron, Chap. 9). For example, Ram and colleagues use a measurement burst design, wherein a single couple provided data for 23 days before the birth of their child and then for 42 days after the birth of their child. This method allowed the researchers to identify daily homeostatic patterns of mood within the couple and to explore how homeostatic processes differed across measurement bursts (i.e., before and after the birth of a child). Moment to moment changes in behavior of 36 mothers and their infants were modeled by Chow and colleagues. The approach allows the capture of dyadic processes that unfold rapidly in time.

Other studies use methodologies that capture changes that unfold over years of time. For example, Teachman (Chap. 1) explores how marital status and race/ethnicity may affect the trajectory of Body Mass Index (BMI) over a 12 year period. Examining change across years yielded findings that marriage, but not cohabitation, is associated with increases in BMI.

Change and Variation

Growth curve models provide exciting opportunities to understand how family factors may influence mean-level changes over time. For example, Teachman (Chap. 1) investigates how individuals change in their body mass over time. In his study, he models the growth of an individual's body mass index (BMI), and subsequently explores how an individual's growth in BMI might be influenced by the development of their spouse's BMI over time. Thus, Teachman provides an excellent example of how family members may influence each other's developmental processes over time.

Daily diary studies further our understanding of family processes by allowing researchers to begin to investigate within-person variability. As Fuligni (Chap. 4) points out, a focus on within-person variability allows researchers to address new research questions. Within person variability refers to the extent to which an individual deviates from his or her own average across the course of time. Of interest is how the deviations from one's "equilibrium" or "set point" are affected by external events and may have implications for individual or family functioning. Daily processes within one couple were explored by Ram and colleagues (Chap. 2) who studied the distinctive set-points in couples' moods and identified predictors of variability around a couple's set point. Within person processes were also studied by Chow and colleagues (Chap. 3), using state space models to capture intra-individual variability for mother and child dyads. A within-person approach allows researchers to move beyond mean-level differences and to understand how variability within an individual may be related to outcomes.

Together these chapters showcase several approaches to addressing change and variation. The authors highlight the reasons why investigators should take care to distinguish different ways of conceptualizing change and variation, and they provide examples of new questions and knowledge that can emerge from this larger portfolio of opportunities. An important caveat is provided by Fuligni (Chap. 4), however. He cautions that the costs of collecting time-intensive within-person or within-family data should be weighed against what knowledge will be gleaned.

Change in Context

The chapters in this volume highlight many ways that researchers can capture change as it unfolds in the broader ecology. Several writers discuss the advantages of gathering data in real time while family members are engaging in interactions in their natural environments. For example, Smyth and Heron (Chap. 9) and Cummings et al., (Chap. 6) argue that a key advantage of using Ecological Momentary Assessment (EMA) data is the ability to capture individuals' experiences in real time. This method may require family members to respond to a set of questions when signaled electronically, in some studies at specific times of the day. With EMA, questions can be answered in everyday settings as family members go about their daily activities without requiring them to recall events or to generalize about their experiences across long periods of time. Thus, EMA satisfies the call for ecological validity (Cummings et al. Chap. 6; Metzler et al., Chap. 12) as well as increases measurement reliability through limiting recall and social desirability biases (Cummings et al., Chap. 6; Smyth and Heron, Chap. 9).

The effects of the environment on individuals may not be universal, but rather may vary based on family characteristics, and several chapters in this volume consider such moderation processes. For example, Smyth and Heron (Chap. 9) review an EMA study showing that levels of family conflict and emotional responsiveness moderate the impact of family hassles on eating disordered behavior: Girls in families that were high in conflict or low in emotional responsivity were more likely to engage in bulimic behaviors on days when they experienced hassles with their families; however, this association was not found for girls from families low in conflict and high in responsivity (Okon et al. 2003). The National Longitudinal Study of

Adolescent Health (Add Health) provides an example of successfully incorporating genetic markers in order to learn how physiological factors interact with environmental influences and affect health in adolescence and young adulthood (Halpern, Harris and Whitsel, Chap. 8). Turning to intervention studies, Duncan (Chap. 15) suggests that researchers examine moderators of intervention effects in order to understand possible heterogeneity in treatment effects. For example, interventions may have their strongest effects on families most in need of the intervention.

Qualitative approaches also advance understanding of the moderators of intervention effects (Weisner, Chap. 10). Using qualitative interview methods, Weisner and his team were able to identify why some families failed to use or only sporadically used the financial benefits and other resources, such as preschool opportunities, that were available to them as part of the New Hope Project. Open-ended interview data assisted researchers in understanding why intervention effects were found for boys but not for girls: Caregivers explained that families of boys were more likely to use New Hope support to place sons in preschools and to more heavily monitor sons given the perception that boys were at greater risk in disadvantaged neighborhoods than girls (Gibson and Weisner 2002). In this way, child gender moderated the effects of the intervention because of the meanings parents attributed to this child characteristic. Other examples of how cultural "meanings" have implications for family dynamics and influences on individual development are provided by Weisner (Chap. 10).

Overcoming Challenges

A third theme that emerges across the chapters in this volume pertains to the development and use of research designs that can overcome some of the challenges inherent in family research. In the previous sections, we discussed the challenges in defining family and measuring family processes that are experienced differently by individual family members—processes and definitions that change across time and vary in their implications across context. Researchers also encounter a multitude of resource and data limitations in conducting family research, from budgetary and other pragmatic constraints to recruitment of multiple family members and participant attrition. Conducting research on families that results in valid and reliable conclusions is challenging. In this section we review research design strategies intended to overcome challenges and advance family research, including: (1) choosing appropriate and efficient study designs, (2) utilizing innovative methods, and (3) balancing tradeoffs.

Appropriate and Efficient Study Designs

Study designs tailored to the researchers' questions can enhance the interpretability of findings and thereby the contribution and reach of a study. Designs chosen for their efficiency also allow researchers to concentrate time and resources on addressing key questions. Below, we discuss two elements that underlie appropriate and efficient research designs: grounding the research and research questions in a strong theoretical or conceptual framework, and relying on valid, reliable measures and established data collection procedures.

"There is nothing so practical as a good theory" (Lewin 1951, p. 169) certainly applies in the study of families. Theory is a strong foundation for appropriate and efficient study designs because it directs attention to whom to study, what to measure, and how to capture analytically the phenomena of interest in efforts to answer a particular research question. For example, in their study of how couple dynamics change after the birth of a child, Ram and colleagues (Chap. 2) drew on family and ecological systems perspectives to choose methods for capturing homeostatic processes. Theory guided the researchers in adapting nonlinear dynamic models to questions about family systems processes, which in turn opened new opportunities for the researchers to test theoretical assumptions on reciprocal relationships that had previously been difficult to study empirically. Theory can also help researchers craft efficient study designs that target particular family processes. Researchers can turn to theories of change to improve the efficiency of study designs for interventions (Duncan, Chap. 15). Developmental theory directs attention to periods of rapid developmental change as a focus for intervention (Duncan and Vandell 2012).

Another critical aspect of choosing an appropriate and efficient study design has to do with the quality of the measures. Interpretations of results are only as valid as the research measures. Scholars should strive to incorporate robust and comprehensive measures. Biomarkers were included, for example, in the Add Health study in an effort to improve the validity and reliability of physical health assessments over previously used self-report measures (Halpern et al., Chap. 8). As another example, using daily measures of family experiences rather than relying on global reports that require participants to generalize about their family experiences over long and often unspecified periods of time is advocated by Smyth and Heron (Chap. 9).

Innovative Methods

Researchers may find traditional methods inadequate for addressing a research question about families. Devising creative alternatives to traditional methods using new ideas and technologies is another way to enhance study designs. Developing innovative methods tailored to the ability and convenience of the participants has many benefits, including the ability to capture a wider range of perspectives, encourage study participation, and improve quality of the data. Designs that include multiple members of the same family as participants, for example, allow researchers to directly capture different perspectives about the "same" family experiences. The meaning of family processes and events can be very different among members of the same family (Amato, Chap. 11). The family experiences of children whose parents are divorced and whose daily lives are spread across two households vary drastically from the family experiences of their parents (Emery, Chap. 7). And, as Emery suggests, "children's psychological life is deeper and richer than what we can capture on useful but limited structured measures," (p. 7), meaning that new measurement approaches may be needed if we are to understand these children's experiences.

Innovative methods are currently being used to capture novel perspectives on the family, and some of them are discussed in this volume. Traditional survey methods are not often designed for use with young children given their verbal and other cognitive demands. The revised MacArthur Story Stem Battery (MSSB) captures the unique perspectives of children by giving them dolls to represent family members and asking the children to act out and narrate stories about their own families (Cummings et al., Chap. 6). The dolls allow children to demonstrate a broad array of family experiences without artificial limits or preconceived ideas. To capture the youth perspectives on family life, Weisner (Chap. 10) gave youths cameras to photograph important people, places, and things in their lives and then asked youths to talk about their photos. Through this innovative data collection method, the researchers encouraged participation and evoked rich narratives from youths that may not have surfaced using traditional methods.

Emerging technologies also create new possibilities for innovative methods. Smart phones are being used for real-time data collection that is ecologically valid and even more convenient for participants than lab based approaches (Smyth and Heron, Chap. 9). For example, researchers can set smart phones to alert participants when to record data in real-time, an approach that can reduce recall bias and missing data, as well as capture family processes within the home environment. Furthermore, participant-friendly methods that remove the burden of remembering when to record data and allow participants to continue with their daily lives have the added benefit of encouraging study participation.

Balancing Tradeoffs

Arguably every research study or program is challenged by limited resources—from material to human. Making careful decisions about which components of the study should be limited and which should be enhanced to best address a particular research question and to make for the most significant research contribution is an essential step in the design of research. Researchers can optimize tradeoffs by balancing the costs and benefits of each component and determining the best compromise.

Strategizes for balancing financial tradeoffs may emerge as a research plan is being developed. When developing the Add Health Study, Halpern et al., (Chap. 8), relied on pilot testing to determine which tradeoffs made the most sense. For example, the researchers originally intended to use an expensive blood pressure monitor that automatically downloaded readings to avoid data bias from manual entry. Pilot tests revealed, however, that these monitors frequently failed, which then required manual entry. For this reason, the researchers chose a less expensive monitor that required manual data entry but provided highly accurate read-outs. The study required purchase of many hundreds of these monitors. Therefore, the pilot test and choice of the less expensive monitor equated to savings of well over $ 100,000.

The multiphase optimization strategy (MOST) can also be used to weigh costs against benefits of a particular research program (Collins, Chap. 14). MOST is an approach to design and implementation that allows researchers to choose components of an intervention program that optimize its benefits and limit its costs. MOST requires researchers to set optimization criteria, which can include funds needed to implement the intervention and the effectiveness of individual intervention components. In determining optimization criteria, Collins emphasizes that researchers should set a reasonable threshold that can realistically be implemented on a larger scale. A more expensive intervention may increase effectiveness but have the disadvantage of limiting the number of families reached by the intervention. Collins explains that researchers can use MOST to isolate the effectiveness of individual intervention components to determine the balance of costs relative to the effectiveness of a component. An optimal tradeoff may involve removing an intervention component if the increase in effectiveness is not large enough to warrant the expense of implementing that component.

Researchers also need to consider the invasiveness of a method when balancing tradeoffs. Noninvasive biomarker data collection methods for Add Health, including measures of weight, height, skin fold thickness, body circumference and blood pressure were purposely chosen (Halpern et al., Chap. 8). The researchers also took additional efforts to secure the confidentiality of the data and gave the participants several options in consenting to the biomarker collections. The flexibility allowed the researchers to still gain partial data among participants who were comfortable with some aspects of the biomarker collection.

Family researchers frequently must balance the tradeoff between richness of measures of family processes and the generalizability of results. Intensive methods measuring within-person variability can allow for rich data but require large investments in time, labor, and participant involvement (Fuligni, Chap. 4). Consequently, these types of studies use smaller samples. Capturing systems dynamics requires intensive measurement, as in Chow and colleagues' (Chap. 3) study of mother-infant interactions. The demands of data collection and analysis (12 raters coding video data on moment-to-moment changes in affective valence of mother and infant behavior) resulted in a sample size of 36 dyads who were able and willing to visit the research laboratory, thus limiting the generalizability of the findings.

Developing a More Complete Picture of Families

The wide range of methodological concerns and strategies covered within the previous chapters of this book highlights how no single study, measure, or method can provide a complete picture of how families operate and how they influence and are influenced by their members. Instead, family scholars must link together the knowledge gleaned from different studies within different disciplines to develop a more complete picture. In this section, we lay out ways that family scholars can work to build a picture of the family.

Linking Micro and Macro Level Perspectives

Family scholarship could benefit from research that bridges micro- and macro- level processes. Research at each level enriches our understanding of families in distinctive and valuable ways, and some examples of cross-level studies are found in this volume. A coding system that measures family conflict on both micro and macro-levels is being developed by Cummings et al, (Chap. 6). The micro-level coding system measures specific constructive and destructive behaviors in family interactions across short time intervals. The macro-level coding system measures the more global components of the interactions, such as overall resolution to the conflict. The researchers use the micro-level coding to understand the interplay between family members and use the macro-level coding to pull out general themes that could be missed on the micro-level. Integrating the two levels build a more complete picture of family interactions and conflicts.

Integrating Qualitative and Quantitative Methods

Integrating qualitative and quantitative methods can create a more complete picture of families. Different methods and measures provide diverse pieces of information on family processes and systems. Integrating multiple measures and methods in a study design can compensate for weaknesses of one approach with the strengths of the other. As a strong proponent of the use of qualitative methods in family research, Weisner (Chap. 10) details how researchers can benefit by integrating qualitative and quantitative methods. For example, researchers can use qualitative evidence to explain quantitative findings and vice versa. As Amato (Chap. 11) states, "qualitative and quantitative approaches illuminate different facets of the social world, and they have much to offer when combined in the same study" (p. 9).

Bridging Disciplines

As the chapters in this book are based on the interdisciplinary Symposium on Family Issues, the authors and editors advocate incorporating frameworks from a range of fields. The broader scholarly community can provide theory and methods to advance family research. The benefits of bridging disciples for developing innovative family research are evident as each section of the volume presents perspectives from diverse disciplines. A model drawn from biology has been used by Ram and colleagues (Chap. 2) to conceptualize and measure families as dynamic systems. Family demographers can apply the latent growth curves models, typically used by developmental psychologists, to model changes in couples' marital happiness over time using longitudinal survey data as demonstrated by Teachman (Chap. 1).

Adaptability

Family studies benefit from methods that are adaptable over a range of contexts. Adaptable methods can provide for comparisons across contexts, and in turn, for a better understanding of family processes. For example, Cummings et al., (Chap. 6), adapted the revised MacArthur Story Stem Battery method so as to measure family dynamics of children of different ages. When children were very young, the researchers used dolls to portray family relationships. As the children aged into adolescence, the dolls were replaced with story stems in an interview format. Because the researchers were able to adapt the method to fit the developmental needs of the participants, they were able to investigate how perspectives on family relationships vary across development stages.

Conducting Research that has Public Health Impact

To achieve impacts on public health, family programs need to be implemented and to prove effective on a population level. Some family programs are too costly to implement with very large groups of participants. In other cases, program evaluation strategies prove too costly. Careful evaluation of family programs is essential for improving public health because these data can be used to make decisions about what programs or program components are worthy of funding.

Developing Efficient and Effective Interventions

Ineffective components of interventions can waste limited resources and reduce the impact and reach of a program. The multiphase optimization strategy (MOST) offers the benefit of creating efficient interventions (Collins, Chap. 14). MOST can be used to determine the effectiveness of individual intervention components and to make decisions about removing less effective intervention components, thereby improving the efficiency of the intervention.

Researchers need to integrate micro-social measures, survey measures, population indicators, and program indicators in family program evaluations given that interventions are implemented on these multiple levels (Metzler et al., Chap. 12). To determine which measures to include in an evaluation, researchers need to take into account the central goals of the program, the available resources, and the implementation of the intervention.

Meaningful Findings

Developing study designs that reduce the risk of spurious findings in intervention research is a goal of Moore and Wood's work (Chap. 13). Specifically, the authors describe approaches to reducing data bias in evaluating interventions for families that can emerge from testing multiple outcomes, as well as from truncation when the program intervention affects the population with reported outcomes differently between the treatment and control groups. Frequently, researchers target numerous outcomes in testing the effectiveness of an intervention; however, increasing the number of outcomes also increases the risk of spurious findings. Limiting outcome measures to a small number most relevant to the intervention mediators is suggested by Moore and Wood. Replication is suggested by Duncan (Chap. 15) as another strategy for reducing the problems of testing multiple outcomes. Research in general can benefit from being replicated across datasets, methods, and subgroups to determine the reliability and generalizability of the results. The experimental design of intervention evaluations provide for cross-site replications.

Suggestions for how researchers can develop study designs to compensate for truncation, when the outcome being measured determines the sample population are detailed by Moore and Wood (Chap. 13). As interventions are designed to affect outcomes only in the treatment group, truncation can bias comparisons between the treatment and control group. For example, Moore and Wood, were concerned that an intervention aimed at improving relationship quality might have the unintended effect of keeping low quality relationships intact in the treatment group, which would make the treatment group appear to have overall lower relationship quality compared to the control group. Two solutions were suggested: researchers can either develop measures that are designed for the entire sample, or treat truncation as a type of sample attrition. These types of strategies help researchers avoid biases common to program evaluations that could otherwise lead to inaccurate interpretations regarding the effectiveness of interventions.

Accessibility of Measures and Findings for Family Programs

Accessibly of measures and study findings is essential if family programs are to have wide reaching impacts. Making measures, data, and findings accessible to both researchers and policy makers increases the scope and success of family programs (Metzler et al., Chap. 12). For example, self-report surveys translated into multiple languages allow for evaluation of programs across different populations. Evaluations also benefit from linking data across multiple sources, such as being able to link a children's hospital and school records to investigate how child health relates to educational achievement. Additionally, for research evaluations to have real-world application, researchers need to make findings accessible to policy makers. User-friendly computerized data report systems that produce easily interpretable graphs so that complex findings can be summarized into straightforward reports are advocated for by Metzler and colleagues.

Future Implications of Innovative Methods

In an effort to provide an overall view of family methods, the chapters in this volume cover a wide spectrum of methods and applications; nevertheless, we would be remiss to imply that a single volume, compiled at one point in time, is capable of covering every innovative method currently available. In the last few decades, the methods available to family scholars have rapidly grown in number and complexity and will likely continue to grow into the future. What implications will this have for family scholars and family research? We argue that scholars can benefit from innovative methods to advance family research, but the increased range and complexity of methods will require changes in how family scholars approach research.

In this concluding chapter we have presented a variety of ways that scholars benefit from using innovative methods in conducting family research. Family scholars can use innovative methods to capture complex family systems and processes. These methods have improved our ability to study families across contexts, through different perspectives, and over varying degrees of time. We can improve the validity of research by using innovative methods that reduce participant attrition and measurement errors. Innovative methods help family scholars parse out causal pathways from complicated family processes that would otherwise be difficult to isolate. As new methods become available, scholars are able to address theories and develop research questions previously outside the scope of traditional methods. From being able to process large amounts of data to collecting biomarkers, innovative methods have significantly increased our means to conduct family research. Family scholarship will undoubtedly benefit from the application of innovative methods, but at what costs?

Family scholars must ensure that research is not blindly driven by methods. Scholars' enthusiasm for innovative methods is not unreasonable given that methods can provide new opportunities to explore previously unreachable questions or to provide a novel perspective. However, scholars must remember to avoid the temptation of using a method solely for its novelty. That is, a method of analysis is not an end in itself but a means to an end. Our primary goals as family researchers are to advance family scholarship and to communicate our findings to the scientific community, policy makers, and the general public. Therefore, pursuit of meaningful questions should be at the center of our research, as should methods that are appropriate to better answer those questions. We should use the most intuitively appealing and comprehensible methods that accurately address those questions for the purpose of illumination, not obfuscation.

Acquiring the skill set needed to master these often complex methods is another prominent concern with embracing innovative methods. As we move forward to more advanced methods, the time investment and training necessary to apply methods could grow substantially. Family scholars will have to devote increasingly more time and intensive training to take advantage of sophisticated methods. Graduate programs may need to include a greater number of course requirements in methods, which could extend the length of time it takes to complete a degree or deter potential scholars who excel in substantive areas but are less proficient in applying complex methods. Postdoctoral scholars may also need to attend more seminars and workshops devoted to keeping up and improving methodological skills. The intensive time commitments needed to learn increasingly sophisticated techniques and remain knowledgeable on emerging methods could have the unintended effect of slowing research productivity or detracting from substantive or theoretical expertise as scholars invest more time in mastering technical aspects of data collection and analysis. The danger of overemphasis on technique represents a potential style over substance dilemma in which we could end up learning more and more about less and less. The goal should be to strike the appropriate balance between substance and technique both in the training and in the conduct of research.

As the diversity of complex methods grows, we also face the dilemma of having more methods at our disposable than any one scholar can hope to master. Researchers will have to make tradeoffs between breadth and depth of understanding methods. For instance, researchers can easily use statistical computer packages to compute complicated equations without ever understanding the specifics behind the equations. This approach increases the breadth of methods at a researcher's disposal but substantially heightens the risk of incorrectly using a method. On the other hand, sacrificing breadth for depth comes with an entirely different set of problems. The more specialized researchers become, the harder it will be to communicate across methodological areas. One possible adaptation to specializations is a rise in research teams that can strategically combine scholars based on complementary skill sets. Academic institutions could also hire (more) methodologists on staff to help scholars correctly implement complicated methods. Additionally, scholars will be increasingly challenged with describing complex methodologies in easy-to-interpret language to ensure their research is accessible to those less familiar with the methodology.

Concluding Remarks

This volume presents the diversity, benefits, and challenges of emerging methods in capturing the complexity of changing family systems and processes. We believe that researchers find it a valuable resource for selecting and developing innovative methods to advance family research. The methods available to researchers are constantly evolving along with changes in family organization and patterns, which promises continuing opportunities for development and innovation in family scholarship. We believe that innovative methods are valuable tools in furthering our understanding of family dynamics, provided that family researchers remain thoughtful in implementing methods best suited for answering meaningful questions.

References

Adamson, L. B., & Frick, J. E. (2003). The still face: A history of a shared experimental paradigm. *Infancy, 4,* 451–473.

Amato, P. R., Booth, A., Johnson, D. R., & Rogers, S. J. (2007). *Alone together: How marriage in America is changing*. Cambridge: Harvard University Press.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge: Harvard University Press.

Cherlin, A. J. (2010). Demographic trends in the United States: A review of research in the 2000s. *Journal of Marriage and Family, 72,* 403–419.

Cox, M. J., & Paley, B. (1997). Families as systems. *Annual Review of Psychology, 48,* 243–267.

Duncan, G., & Vandell, D. (2012). *Understanding variation in the impacts of human capital interventions on children and youth, Irvine Network on individuals in development*. Working paper.

Elder, G. H. (1998). The life course and human development. In R. M. Lerner, & W. Damon (Eds.), *Handbook of child psychology, Vol. 1. Theoretical models of human development* (pp. 939–991). New York: Wiley.

Gibson, C., & Weisner, T. S. (2002). Rational and ecocultural circumstances of program take-up among low-income working parents. *Human Organization, 61*(2), 154–166.

Lewin, K. (1951). *Field theory in social science: Selected theoretical papers by Kurt Lewin*. London: Tavistock.

Lippold, M. A., Greenberg, M. T., & Feinberg, M. E. (2011). A dyadic approach to understanding the relationship of maternal knowledge of youths's activities to youths's problem behavior among rural adolescents. *The Journal of Youth and Adolescence, 40,* 1178–1191.

Minuchin, P. (1985). Families and individual development: Provocations from the field of family therapy. *Child Development, 56,* 289–302.

Okon, D. M., Greene, A. L., & Smith, J. E. (2003). Family interactions predict intraindividual symptom variation for adolescents with bulimia. *International Journal of Eating Disorders, 34,* 450–457.

Repetti, R., Wang, M. A., & Saxbe, D. E. (2011). Adult health in the context of everyday family life. *Annals of Behavioral Medicine, 42,* 285–293.

Smetana, J. G. (1988). Concepts of self and social convention: Adolescents's and parents's reasoning about hypothetical and actual family conflicts. In M. R. Gunnar & W. A. Collins (Eds.), *Development during the transition to adolescence. The Minnesota Symposia on Child Psychology, Vol. 21*, (pp. 79–122). Hillsdale: Lawrence Erlbaum & Associates.

Thornton, A., Axinn, W. G., & Xie, Y. (2007). *Marriage and cohabitation*. Chicago: University of Chicago Press.

Tronick, H., Adamson, L., Wise, S., & Brazelton, T. B. (1978). The infant's response to entrapment between contradictory messages in face-to-face interaction. *American Academy of Child Psychiatry, 17,* 1–13.

Index

S. M. McHale et al. (eds.), *Emerging Methods in Family Research,*
National Symposium on Family Issues 4, DOI 10.1007/978-3-319-01562-0,