Edited by
Luna C. Centifanti and David M. Williams

THE WILEY HANDBOOK OF *Developmental Psychopathology*

WILEY Blackwell

The Wiley Handbook of Developmental Psychopathology

The Wiley Handbook of Developmental Psychopathology

Edited by

Luna C. Centifanti
David M. Williams

WILEY Blackwell

This edition first published 2017

Registered Offices
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Office
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data

Names: Centifanti, Luna C., 1974– editor. | Williams, David M., 1980– editor.
Title: The Wiley handbook of developmental psychopathology / [edited by] Luna C. Centifanti, David M. Williams.
Description: Hoboken, NJ : John Wiley & Sons, 2017. | Includes bibliographical references and index.
Identifiers: LCCN 2016055391 (print) | LCCN 2017001686 (ebook) | ISBN 9781118554555 (cloth) | ISBN 9781118554548 (pdf) | ISBN 9781118554531 (epub)
Subjects: LCSH: Psychology, Pathological–Handbooks, manuals, etc. | Developmental psychology–Handbooks, manuals, etc.
Classification: LCC RC454.4 .W544 2017 (print) | LCC RC454.4 (ebook) | DDC 616.89–dc23
LC record available at https://lccn.loc.gov/2016055391

Cover Image: © alepvfoto/Fotolia
Cover Design: Wiley

Set in 10/12pt Galliard by SPi Global, Pondicherry, India

10 9 8 7 6 5 4 3 2 1

Contents

List of Contributors

Nathalie M. G. Fontaine, School of Criminology, University of Montreal

Isaac T. Petersen, Department of Psychological and Brain Sciences, Indiana University

Henrik Larsson, Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden

Brian M D'Onofrio, Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

Madeleine J. George, Duke University

Michael A. Russell, Penn State University

Candice L. Odgers, Duke University

Nicola Shaughnessy, School of Arts, University of Kent

Cynthia S. Peng, Department of Speech, Language, and Hearing Sciences, George Washington University, USA

Gregory L. Wallace, Department of Speech, Language, and Hearing Sciences, George Washington University, USA

Nicola Botting, Division of Language and Communication Science, School of Health Sciences, City, University of London

Chloë Marshall, UCL Institute of Education, University College London

Jarla Pijper, Adolescent Development, Utrecht University & Clinical Child and Adolescent Studies, Leiden University

Minet de Wied, Adolescent Development, Utrecht University

Stephanie van Goozen, Clinical Child and Adolescent Studies, Leiden University & School of Psychology, Cardiff University

Wim H. J. Meeus, Adolescent Development, Utrecht University & Developmental Psychology, Tilburg University

Stefanie Hoehl, 1. Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany; 2. Department of Applied Psychology: Health, Development, Enhancement and Intervention, University of Vienna

Emily J.H. Jones, Brain and Cognitive Development, Birkbeck, University of London

Mark H. Johnson, Brain and Cognitive Development, Birkbeck, University of London

Dr Alice Jones Bartoli, University of London

Prof Alice M. Gregory, Goldsmiths, University of London

Kostas A. Papageorgiou, 1. School of Psychology, Queen's University Belfast, UK; 2. Department of Psychology, Tomsk State University, Russia

Angelica Ronald, Department of Psychological Sciences, Birkbeck, University of London

Yulia Kovas, (1) Goldsmiths, University of London; (2) Tomsk State University

Maria Grazia Tosto, Tomsk State University

David M. Williams, School of Psychology, University of Kent

David J. Hawes, PhD, School of Psychology, The University of Sydney

Luna C. Muñoz Centifanti, University of Liverpool, UK

Scott Risser, Montana Tech, USA

Bethany Little, Durham University, UK & Vrije Universiteit, Netherlands

Hedwig Eisenbarth, University of Southampton

Dale M. Stack, Concordia University, Department of Psychology and Centre for Research in Human Development

Lisa A. Serbin, Concordia University, Department of Psychology and Centre for Research in Human Development

Célia Matte-Gagné, Concordia University, Department of Psychology and Centre for Research in Human Development

Danielle Kingdon, Concordia University, Department of Psychology and Centre for Research in Human Development

Kelly Doiron, Concordia University, Department of Psychology and Centre for Research in Human Development

Alex E. Schwartzman, Concordia University, Department of Psychology and Centre for Research in Human Development

Amanda Sheffield Morris, Oklahoma State University

Benjamin J. Houltberg, Fuller Theological Seminary

Michael M. Criss, Oklahoma State University

Cara D. Bosler, Oklahoma State University

Carl F. Weems, Iowa State University

Brandon G. Scott, Arizona State University

Jacob A. Burack, Department of Educational and Counselling Psychology, McGill University, Montreal, Quebec, Canada

Ashley Reynolds, Department of Psychology, Concordia University, Montreal, Quebec, Canada

Oriane Landry, School of Psychology and Public Health, La Trobe University Melbourne, VIC, Australia

Gillian Klassen, Department of Educational and Counselling Psychology, McGill University, Montreal, Quebec, Canada

Natalie Russo, Department of Psychology, Syracuse University, Syracuse, New York, USA

Stephanie A. Fryberg, Department of Psychology and American Indian Studies, University of Washington, Seattle, Washington, USA

Valsamma Eapen, School of Psychiatry, Faculty of Medicine, University of New South Wales, Australia

Rudi Črnčec, School of Psychiatry, Faculty of Medicine, University of New South Wales, Australia

William Mandy Department of Clinical, Educational and Health Psychology, Division of Psychology & Language Sciences, Faculty of Brain Sciences, University College London

Paul J. Frick, Louisiana State University & Australian Catholic University

Laura C. Thornton, Boys Town, Omaha, Nebraska USA

Kathryn L. Modecki, Griffith University, Brisbane, Australia

Bep Norma Uink, Murdoch University, Western Australia

Section I

Developmental Psychopathology: An up-to-date Historical and Methodological Overview

Part 1

Methods for Studying Developmental Psychopathology

1

Developmental Trajectories of Psychopathology

An Overview of Approaches and Applications

Nathalie M. G. Fontaine and Isaac T. Petersen

Introduction

Developmental psychopathology concerns the study of the development of psychological disorders (e.g., depression, anxiety, schizophrenia, conduct problems), risk and protective factors, as well as outcomes, through a lifecourse perspective (Cicchetti, 1989; Rutter, 1990). Longitudinal studies, which involve repeated measures of the same variables from the same individuals, are crucial to investigate change (increases or decreases), but also stability, of psychopathology over time. Indeed, unlike cross-sectional studies, in which different individuals are compared at one time point, longitudinal studies allow for (1) the exploration of within-individual change (or stability)—that is, how each individual develops over time—and (2) between individual differences—including the investigation of distinct patterns of change (or stability) over time across individuals and factors associated with these distinct patterns (Singer & Willett, 2003).

More specifically, longitudinal data enable testing hypotheses about the development of behaviors, the developmental association between different, yet related behaviors (e.g., hyperactivity and physical aggression) and the factors associated with stability or change of behaviors over time. The identification of factors associated with persistence or change in behavior (increasing or decreasing patterns) can shed light on the vulnerabilities associated with severe and persistent psychopathology. In turn, a better understanding of risk processes in the development of psychopathology can help in the development of effective intervention strategies that target protective factors associated with desistance or decreased levels of psychopathology. Longitudinal data can therefore be useful for testing developmental theories about psychopathology as well as the effectiveness of prevention and treatment programs (Cicchetti & Toth, 1992).

New advances in statistical approaches over the last decades help in maximizing what we can learn using longitudinal data in the field of psychopathology (Nagin, 2005). Methodologists have developed various statistical approaches, which include and are known variously as growth curve models (GCMs), random coefficient models, multilevel models, mixed models, mixed-effects models, hierarchical linear models,

The Wiley Handbook of Developmental Psychopathology, First Edition. Edited by Luna C. Centifanti and David M. Williams.

group-based trajectory models (GBTMs), latent class growth models (LCGMs), and growth mixture models (GMMs) (Muthén, 2004; Nagin, 2005; Singer & Willett, 2003). The approaches have been applied with a rapid rise in many areas of developmental psychopathology research (Curran, Obeidat, & Losardo, 2010; Nagin & Odgers, 2010), including conduct problems (Nagin & Tremblay, 1999), depression/anxiety (Côté et al., 2009), callous-unemotional traits (Fontaine, Rijsdijk, McCrory, & Viding, 2010), and substance use problems (Hu et al., 2008). These approaches have also been applied to assess heterogeneity in treatment responses to clinical trials (Muthén et al., 2002).

Statistical approaches for longitudinal data can be complex in terms of selecting the optimal approach, fitting the models to the data and interpreting the findings with respect to hypothesis and theory (Curran et al., 2010). In this chapter, we introduce some of the approaches and applications, particularly to non-technical readers, including researchers, clinicians, and graduate students, who may not have yet received an extensive training in this area. References to more detailed and complete technical developments on these approaches are offered for interested readers. We first provide an overview of the approaches, with a focus on GCM, GBTM and GMM, namely approaches focusing on developmental trajectories (Muthén, 2004; Nagin & Odgers, 2010). Next, we present selected examples of applications of these models in the field of developmental psychopathology and clinical psychology. Finally, we discuss methodological considerations when applying these models and interpreting the findings.

Overview of the Approaches

This section presents an overview of three approaches applied to longitudinal data, and more specifically to trajectory modeling: GCM, GBTM and GMM. We selected these three approaches because they share a common analytical goal, namely to examine differences or variability across members of a population in their developmental trajectories (Nagin & Odgers, 2010). A developmental trajectory can be defined as the course of a behavior over time or age (Nagin, 1999). Because these trajectory analyses examine longitudinal data with more than two measurement occasions, they have key advantages over analytical approaches that do not examine trajectories (Beauchaine, Webster-Stratton, & Reid, 2005; Gueorguieva & Krystal, 2004). First, trajectory analyses tend to have better reliability and greater power to detect behavioral change than simple pre–post or difference score designs. Second, trajectory analyses have greater flexibility with unbalanced designs, unequal spacing of time points, and tolerate missing data, unlike repeated measures analysis of variance (ANOVA). Third, trajectory analyses are less likely to inflate the Type I error rate than are repeated measures ANOVA analyses, which have more strict assumptions (e.g., sphericity). Fourth, trajectory analyses often allow multiple outcomes to be examined in the same analysis. Although these approaches share a common goal, they make distinct assumptions about the distribution of trajectories in the population. Figure 1.1 presents hypothetical trajectories according to the GCM, GBTM, and the GMM approaches. In a nutshell, it is assumed with GCM that all individuals come from the same population and can be described by the same parameters of change. It is not assumed, however, that individuals' change is identical—the model captures the average developmental trend and person-specific variations around the average trend using the same parameters of

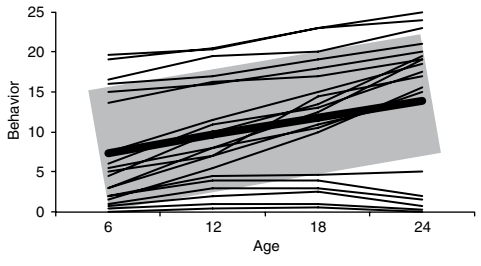
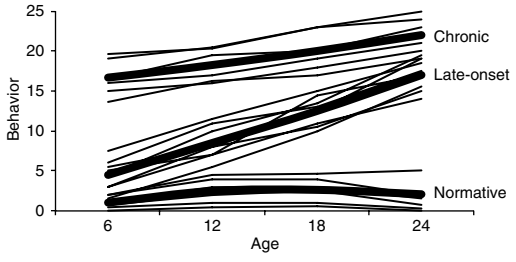
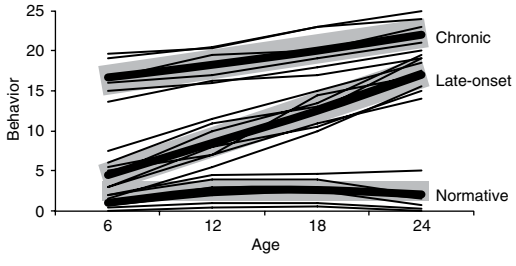
Approach	Trajectories	Description
GCM		Assumes that all individuals come from the same population, but allows individuals' change to differ—that is, the model captures the variations that are specific to each individual (random effects). The thin black lines denote the observed data points, the thick solid black line denotes the average trajectory of change and the shaded gray area around the line denotes the individual variations (Shiyko et al., 2012).
GBTM		Assumes that some individuals come from distinct subpopulations, as captured by different subgroups. The subgroups are assumed to be homogenous regarding change—that is, that all individuals on a particular trajectory have the same pattern of behavior over time. The thin black lines denote the observed data points. For each trajectory, the thick solid black line denotes the average trajectory of change (Shiyko et al., 2012).
GMM		Assumes that some individuals come from distinct subpopulations, as captured by different subgroups. However, contrary to GBTM, heterogeneity within each subgroup is also considered in GMM—that is, variations that are specific to each individual (random effects) like in GCM can also be modeled. The thin black lines denote the observed data points. For each trajectory, the thick solid black line denotes the average trajectory of change and the shaded gray area around the line denotes the individual variations (Shiyko et al., 2012).

Figure 1.1 Hypothetical Trajectories According to the GCM, GBTM, and the GMM Approaches.

change. GBTM and GMM, by contrast to GCM, assume that some individuals come from distinct subpopulations, as captured by different subgroups. These subgroups can be described by different parameters of change. In other words, the models allow different individuals to follow different trajectories, but only GBTM and GMM allow

Table 1.1 Summary of the key outputs of the GCM, GBTM, and the GMM approaches.

	GCM	*GBTM*	*GMM*
Intercept and slope of development	✔	✔ (for each trajectory)	✔ (for each trajectory)
Individual-specific effects (random effects)	✔		✔ (for each trajectory)
Proportion of the population following each developmental trajectory		✔	✔

subgroups (that are not captured by model predictors) of individuals with qualitatively different forms of change.

A number of differences exist between GCM, GBTM and GMM approaches. Researchers in developmental psychopathology often have to decide what approach they should apply. We present below a brief summary of the assumptions underlying each approach to help readers decide the optimal strategy for a given research question or hypothesis. Table 1.1 presents a summary of the key outputs of the three different approaches.

GCM

In their simplest form, GCMs typically fit a best-fit straight line to each individual's trajectory of change over time. Each individual's line is allowed to have a different starting point (intercept) and direction and steepness of change (slope). Each individual's best-fit line is slightly adjusted to take into account the trajectories of the other individuals in the sample—a phenomenon known as shrinkage because individuals' GCM estimates are shrunk towards the mean estimate for the sample, making the GCM estimates more reliable (i.e., having less measurement error) than estimates from simple regression (Hox, 2010). Based on theory and/or how well the model fits the data, the modeler can decide whether each parameter (intercept, slope) is the same (fixed effect) or allowed to differ (random effect) between individuals. GCM can be extended to consider nonlinear forms of change, such as polynomial (e.g., quadratic), exponential, and logistic forms.

GCM can be fit in a structural equation modeling (SEM) or hierarchical linear modeling (HLM; also known as multilevel modeling, mixed modeling, or mixed-effects modeling; Raudenbush & Bryk, 2002) framework. In general, SEM is more advanced and flexible than HLM. SEM, unlike HLM, allows specifying latent variables that represent the common variance among observed (manifest) variables, and have less measurement error. Unlike HLM, SEM also allows specifying multiple outcomes in the same analysis and more flexibility in specifying correlated residuals (which, if residuals covary, would violate assumptions if unspecified). However, SEM typically requires a larger sample size than does HLM. In addition, HLM is more flexible when participants are not sampled at the same time points because HLM uses data in long form (rather than wide form in SEM). In long form, each row represents a combination of participant and measurement occasion (i.e., 100 participants × 3 measurement occasions = 300 rows, 100 variables = 100 columns). In wide form, each row represents one participant, with columns representing combinations of variable and measurement occasion (i.e., 100 participants = 100 rows, 100 variables × 3

Table 1.2 Example of data in wide form and in long form. Data on externalizing problems in wide and long forms for three participants and four time points.

Wide form						*Long form*			
ID	*Sex*	*Ext_1*	*Ext_2*	*Ext_3*	*Ext_4*	*ID*	*Sex*	*Time*	*Ext*
1	F	1	0	2	Missing	1	F	1	1
2	F	2	6	8	Missing	1	F	2	0
3	M	4	7	6	8	1	F	3	2
						2	F	1	2
						2	F	2	6
						2	F	3	8
						3	M	1	4
						3	M	2	7
						3	M	3	6
						3	M	4	8

Abbreviations: ID = identification number; Ext = externalizing problems.

measurement occasions = 300 columns). If participants are not sampled at the same time points, however, wide form can be computationally cumbersome because it can result in too many columns where most participants have missing values (see Table 1.2 for an example of data in wide form and in long form). HLM (which uses long form—a more efficient data structure with less missingness when participants are sampled at different time points) can be advantageous in these circumstances. See Curran (2003) and Bauer (2003) for further discussion of similarities and differences between the two approaches, and when to use each.

GCM assumes that all individuals in the sample come from the same population and can be described by the same parameters of change (e.g., everyone could be described with a quadratic trajectory—a U-shaped trajectory—in which scores of psychopathology would first decrease and then increase over time). GCM does not necessarily assume, however, that all individuals show identical change. For example, a polynomial GCM can allow each individual to have different intercepts, slopes, and curvatures. GCM grasps the average developmental trend and person-specific variation around the average trend (random effect) using the same parameters of change for the population. Furthermore, GCM can also allow subgroups (e.g., males and females) with different trajectories, but the subgroups have to be described by the same parameters of change. Thus, individual or subgroup differences in GCM trajectories are assumed to reflect quantitative differences in degree rather than qualitative differences in kind. To examine qualitative differences between subgroups of individuals that are not captured by model predictors, however, one must use another type of model such as GBTM or GMM.

GBTM

GBTM has been developed by Nagin and colleagues (Nagin, 1999, 2005; Nagin & Tremblay, 1999). With this approach, the population under study is considered to be composed of qualitatively distinct subgroups that are not detectable using measured

characteristics (Nagin & Odgers, 2010). For instance, in a study on depressive symptoms, two different subgroups of individuals could compose the population, hypothetically one subgroup with a genetic vulnerability for depression and one without the vulnerability. If the subgroups are not distinguishable based on measured characteristics, we could assume that the dataset is composed of a combination of two different subgroups (Nagin & Odgers, 2010). Using GBTM, distinct statistical processes would be involved to explain the development of depressive symptoms in the two subgroups. Owing to distinct etiologies, the different subgroups might show different trajectories, which is the qualitative heterogeneity we seek to capture with subgroups in GBTM (as opposed to quantitative heterogeneity in GCM). The distinct subgroups are assumed to be homogeneous regarding change—that is, that all individuals on a particular trajectory have the same pattern of psychopathology (e.g., depressive symptoms) over time. Whether or not the trajectories really differ qualitatively as hypothesized may be subject to interpretation (Bauer & Reyes, 2010). In some instances, the trajectories may reflect points on a continuum as if a continuous variable was divided into different ordered-categorical levels (e.g., low levels, moderate levels and high levels of psychopathological symptoms). Still, the subgroups' trajectories could be distinguished by different predictors (e.g., childhood maltreatment) or be associated with different outcomes (e.g., job dissatisfaction).

Key outputs of GBTM are the proportion of the population following each trajectory and each individual's probability of being classified in each trajectory (probability of membership in each trajectory). GBTM takes into account uncertainty in group assignment. Table 1.3 presents, as an example, hypothetical scores of conduct problems between ages 6 to 12 for four participants. Based on their scores and assuming that the GBTM estimated four trajectories (i.e., stable high, increasing, decreasing and stable low), the model should most likely classify with a high probability (e.g., 0.75 and above) Sam as following a stable high trajectory, Gabby as following an increasing trajectory, and Max as following a decreasing trajectory. Fred's assignment to a group may be more challenging. For instance, Fred may have 0.51 probability to follow the stable high trajectory, 0.20 probability to follow the stable low trajectory, 0.10 probability to follow the increasing trajectory, and 0.19 probability to follow the decreasing trajectory (the probabilities sum to 1). Based on the probabilities, Fred would be classified in the stable high trajectory. Although using the trajectory membership (i.e., the assignment to trajectories according to the probabilities of each individual's most likely group membership) can be clinically relevant, it is important to consider that the trajectories are most likely approximations of a more complex reality (Nagin & Odgers, 2010).

Table 1.3 Hypothetical scores of conduct problems from 6 to 12 years old for four participants.

	Scores of conduct problems between 6 to 12 years old						
Participants	*6*	*7*	*8*	*9*	*10*	*11*	*12*
Sam	9	8	6	10	8	6	7
Gabby	2	4	5	4	6	8	10
Max	10	7	7	6	4	2	1
Fred	7	2	6	3	8	4	6

GMM

GMM has been developed by Muthén and colleagues (Muthén & Shedden, 1999). This approach allows both qualitatively different patterns of change and quantitative differences within each pattern (Bauer & Reyes, 2010). Similarly to GBTM, the population under study is assumed to be composed of qualitatively distinct subgroups. However, heterogeneity within each subgroup is also considered—that is, variations that are specific to each individual (random effects) as in GCM can also be modeled within each subgroup. In GMM, the subgroups could not be explained by the between-individual variability accounted for by individual-specific effects in a single GCM (random effects). Instead, different single GCMs capture the population variability across time. In GMM, two or more GCMs are used to estimate the population variability in developmental trajectories. Key outputs of such an approach are two or more GCMs and estimates of the proportion of the population following each GCM.

Both GBTM and GMM provide helpful information about the shape and levels of the trajectories. The main difference between the two approaches is that GMM includes random effects (i.e., individual-specific effects) in each trajectory, whereas GBTM does not. Adding random effects can lead to a fewer number of trajectories because it allows for more within-group variability in individual-level trajectory (Nagin & Odgers, 2010).

Examples of Applications

Studying Externalizing and Internalizing Problems Using GCM

In the following section, we present examples of how GCM can be used to study externalizing and internalizing problems across ages. We discuss linear (i.e., linear trend over time) and nonlinear (e.g., quadratic trend over time) models.

Linear Growth Curves The most common GCMs fit linear growth curves. Owens and Shaw (2003) examined linear trajectories of mother-reported externalizing problems of children (n = 299) from 2 to 6 years of age from low-income families. On average, children showed decreases in externalizing problems over time. Children with more externalizing problems at age 6 were characterized by more negative emotionality as infants and having depressed mothers, mothers who were less accepting of the child's behavior, and parents with more inter-parental conflict. Keiley, Bates, Dodge, and Pettit (2000) used SEM to simultaneously examine linear growth curves of mothers' and teachers' reports of children's (n = 405) internalizing and externalizing problems from 5 to 12 years of age. Demonstrating the developmental comorbidity between internalizing and externalizing problems, intercepts and slopes of internalizing problems were positively correlated with the intercepts and slopes, respectively, of externalizing problems. In other words, the higher one's initial level of internalizing problems, the higher one's initial level of externalizing problems (and the greater one's increases in internalizing problems, the greater one's increases in externalizing problems). Children with higher intercepts of externalizing problems included children rejected by peers, children from lower socioeconomic status (SES) families, and European American (compared to African American) children. Children with greater increases of externalizing problems included boys, rejected children, and African American (compared to European American) children. In summary, GCMs

can test the effect of risk and protective factors on intercepts (initial level) and slopes (growth/change) of developmental trajectories, and can fit multiple behaviors simultaneously to understand their codevelopment.

Nonlinear Growth Curves In addition to linear growth curves, previous studies using GCMs have examined nonlinear growth curves. Studies have examined polynomial trajectories of externalizing problems, including quadratic (e.g., Bongers, Koot, van der Ende, & Verhulst, 2003), cubic (e.g., Nærde, Ogden, Janson, & Zachrisson, 2014), and quartic (e.g., Petersen, Bates, Dodge, Lansford, & Pettit, 2014) trajectories. Bongers and colleagues (2003) examined quadratic trajectories of parent-reported externalizing problems (in addition to other behavior problems) in 2,076 Dutch children from ages 4 to 18 years. The sample showed curvilinear decreases in externalizing problems over time, on average. Boys had higher intercepts and steeper slopes (greater decreases) than girls, yet were reported to show more externalizing problems than girls across time. Nærde and colleagues (2014) examined cubic trajectories of parent-reported physical aggression in 1,159 Norwegian children from 8 to 26 months of age. Aggression increased from 8 months, with a peak around 20 to 22 months, followed by a decrease to 26 months. Higher intercepts were found among boys (compared to girls), those having a same-age sibling, higher levels of parental distress, lower parental education, more difficult child temperament, and higher levels of child activity. Slopes of aggression were predicted by the presence of a same-age sibling and high child-activity level. It is worth noting that it can be quite difficult to interpret the predictors of polynomial terms (Grimm, Ram, & Hamagami, 2011). In any case, GCMs can fit nonlinear trajectories and estimate the effects of risk and protective factors on the level and growth of these trajectories.

Joint Development of Related yet Different Behaviors Using GBTM

In this section, we present an example of trajectory modeling of two related yet different behaviors. This type of model (referred to as a dual or joint model) allows examining the developmental relations between two behaviors of interest (Nagin, 2005; Nagin & Tremblay, 2001). The joint model has three key outputs: (1) the developmental trajectories for each type of behavior; (2) the probability of membership in each joint trajectory (which reflects the proportion of individuals belonging simultaneously to trajectories of both types of behaviors (e.g., being classified in the high trajectory of one type of behavior and in the low trajectory of the second type of behavior); and (3) the probabilities linking membership in trajectories across behaviors (e.g., the probability of following a high trajectory of one type of behavior given the trajectory membership of the other behavior).

In a study of girls from a population-based sample (n = 881), Fontaine and colleagues (2008) examined the joint developmental trajectories of physical aggression and hyperactivity between 6 to 12 years old (teachers assessed the behaviors yearly) using GBTM. They identified four trajectories of hyperactivity (None/Low, Moderate declining, Moderate stable and High declining) and three trajectories of physical aggression (None/Low, Moderate, and High declining) (first key output). They also found that about 10% of the girls belonged to the high trajectory of hyperactivity without high levels of physical aggression, that a small proportion of the girls (less than 1% of the sample) belonged to the high trajectory of physical aggression without high levels of hyperactivity, that about 9% of the girls belonged to the high

trajectory of hyperactivity *and* the high trajectory of physical aggression, and that the remainder of the sample encompassed all other trajectory combinations (second key output). Finally, they found an asymmetric developmental association between hyperactivity and physical aggression; in other words, girls with high levels of physical aggression were highly likely to have high levels of hyperactivity (probability = 0.96), but girls with high levels of hyperactivity were only moderately likely to have high levels of physical aggression (probability = 0.45). In addition, girls who were not hyperactive were likely not to be physically aggressive (probability = 0.91), but girls who were not physically aggressive could show moderate levels of hyperactivity (probability = 0.28 for moderate declining levels and probability = 0.20 for moderate stable levels). Figure 1.2 illustrates

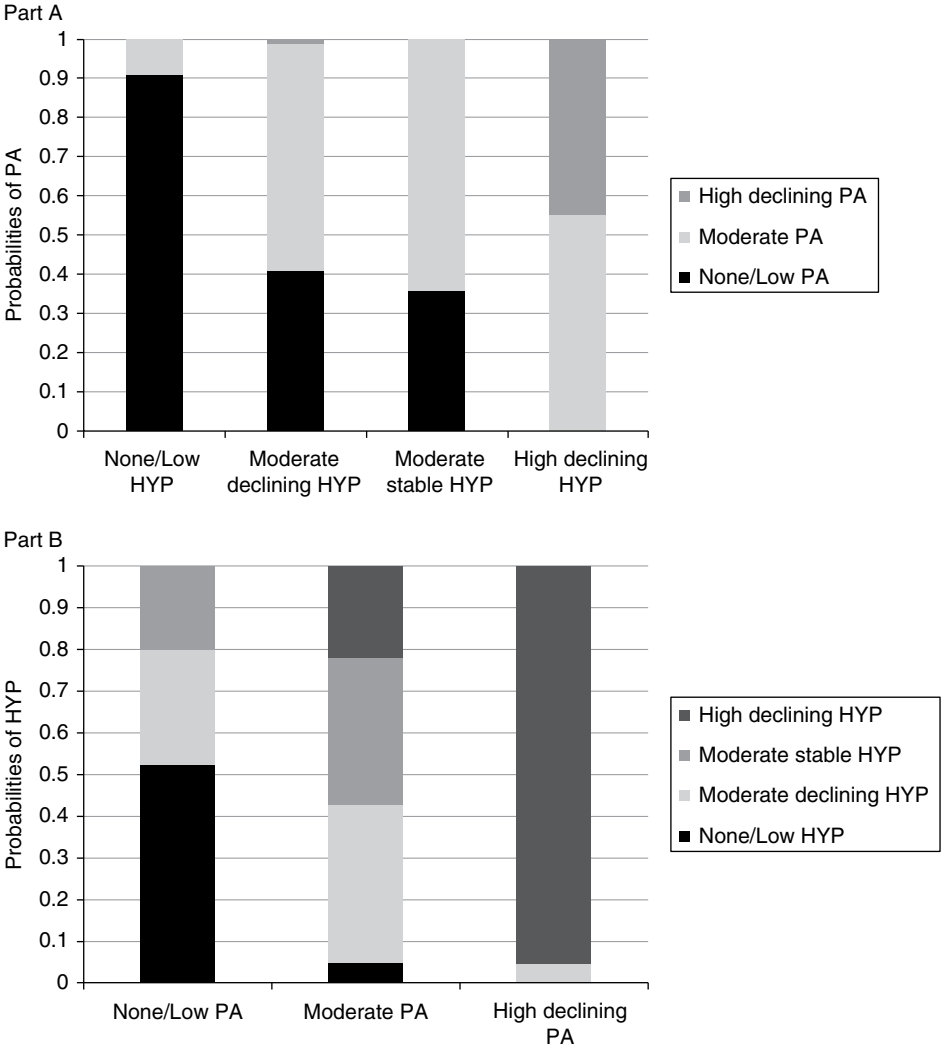


Figure 1.2 Linking Probabilities between the Trajectories of Hyperactivity and Physical Aggression (Based on Fontaine et al., 2008). Part A: Probabilities of Physical Aggression (PA) Conditional on Hyperactivity (HYP); Part B: Probabilities of Hyperactivity (HYP) Conditional on Physical Aggression (PA).

these linking probabilities (third key output). In sum, GBTM allows for the study of the joint development of related yet different behaviors, including the probabilities linking membership in trajectories across behaviors.

Testing Developmental Taxonomical Theories or Theoretical Models

Trajectory modeling can be useful to test taxonomical theories or theoretical models. In this section, we present studies that focused on testing theoretical models, such as Moffitt's taxonomical theory of antisocial behavior.

The Example of Moffitt's Developmental Taxonomy Moffitt (1993) postulated that antisocial behavior can follow two distinct trajectories, each with a unique etiology and prognostic: a small group of individuals engaging in antisocial behavior at an early age and persistent throughout their lifespan (lifecourse-persistent trajectory) and a larger group of individuals engaging in antisocial behavior only during adolescence (adolescence-limited trajectory). According to her theory, the interaction between childhood neuropsychological problems and adverse environments could lead to lifecourse-persistent antisocial behavior, whereas a contemporary maturity gap and involvement with deviant peers could lead to adolescence-limited antisocial behavior.

Odgers and colleagues (2008) used GMM to test Moffitt's (1993) taxonomic predictions regarding developmental course of antisocial behavior, risk factors and adult consequences. They identified four antisocial behavior trajectories: lifecourse-persistent, adolescence-onset, childhood-limited and low. In line with Moffitt's theory, their findings supported the existence of both lifecourse-persistent and adolescence-onset antisocial trajectories. However, contrary to the taxonomy, some individuals were identified with an adolescence-onset antisocial behavior, but also subsequent persistence into adulthood. Although Moffitt's theory predicted that variability in age at desistence could be explained by the cumulative number and type of ensnaring life events that entangle individuals in an antisocial lifestyle, methodological issues could also explain the findings by Odgers and colleagues (e.g., spacing of measurement occasions during adolescence and the challenges of mapping such transient behavioral pattern using trajectory modeling). In addition, a childhood-limited trajectory, not anticipated by the taxonomy, was identified. The individuals engaging in the childhood-limited trajectory appeared to desist from antisocial behavior, however, they were likely to experience small to moderate problems that were restricted to internalizing disorders, smoking and financial difficulties.

The examples above demonstrate that trajectory modeling has implications for testing developmental theories, which could at the same time inform research and clinical practice. For instance, given the different developmental course of each trajectory, as well as their associated risk factors and outcomes, targeted interventions that best address the profiles of vulnerabilities and strengths of individuals following distinct trajectories could be developed (Viding, Fontaine, & McCrory, 2012).

A Dimensional Alternative to Moffitt's Developmental Taxonomy There is a growing body of research showing that externalizing problems are dimensional rather than categorical in nature (Coghill & Sonuga-Barke, 2012; Krueger, Markon, Patrick, & Iacono, 2005; Markon & Krueger, 2005; Walton, Ormel, & Krueger, 2011). Recent findings also suggest that trajectories of externalizing problems are more accurately

modeled dimensionally than with Moffitt's (1993) subgroups (Walters, 2011, 2012; Walters & Ruscio, 2013; for a review, see Fairchild, van Goozen, Calder, & Goodyer, 2013). The dimensional nature of the development of externalizing problems suggests that GCM may be useful when modeling trajectories of externalizing problems because GCM seeks to capture quantitative variation in individuals' trajectories.

However, one difficulty in measuring change in externalizing problems across long developmental periods is that the meaning of externalizing behaviors changes with development. For instance, physical aggression is more common in young children than adolescents (Miller, Vaillancourt, & Boyle, 2009), so physical aggression may likely represent a more severe trait level of externalizing behaviors in adolescence than it does in early childhood. Because the meaning of externalizing behaviors changes with development (known as heterotypic continuity), the same measure may have a different meaning at different ages. Thus, developmental changes in the construct of externalizing behavior may necessitate changes in measurement. A widely used measurement system for externalizing behavior, the Achenbach System of Empirically Based Assessment (ASEBA; Achenbach, 2009) accommodates changes in the construct of externalizing behaviors by including different items at different ages in development. A key challenge is ensuring that differences across time reflect actual change rather than differences in the meaning of the measure. One approach to make scores on different measures statistically comparable is to use a proportion score (Little, 2013), where the child's problem sum is divided by the total possible score on the measure.

Petersen and colleagues (2015) examined GCMs of externalizing problems from 5 to 27 years of age ($n = 585$) using proportion scores of ratings by mothers, fathers, teachers, peers and self-report on the ASEBA. Quartic (fourth-degree polynomial) trajectories were the best-fitting form of change. On average, externalizing problems decreased from early childhood to preadolescence (ages 5–11), increased during adolescence (11–16), and decreased from late adolescence to adulthood (16–27). Yet there were great individual differences in intercepts and slopes that were predicted by risk factors reflecting family process, peer process, stress, and individual characteristics. The quartic model with these risk factors as predictors was fairly accurate, and accounted for 70% of the variability in the development of externalizing problems. In sum, GCMs can accurately capture dimensional individual differences in the development of externalizing problems over long developmental spans with changing measures.

Testing the Effectiveness of Interventions Using Trajectory Modeling

Trajectory modeling can be applied to assess the effects of intervention strategies nested in longitudinal studies, such as longitudinal randomized trials. In the following section, we present studies in which GBTM, GMM and GCM were applied to examine interventions' effects in longitudinal randomized trials.

GBTM and GMM to Test the Effectiveness of Randomized Preventive Interventions GBTM and GMM can be used to examine the impact of interventions on subgroups characterized by different growth trajectories, that is the effects on populations that include individuals who have normative and non-normative patterns. The strength of such analysis is that it allows the assessment of intervention effects on the different trajectories rather than focusing on overall intervention effects at a specific time point.

Lacourse and colleagues (2002) used GBTM to assess the effects of a randomized multimodal preventive intervention in boys (n = 909). Boys with high levels of disruptive behavior (based on the teachers' assessment in kindergarten) were randomly assigned to the intervention group (n = 42) or the control group (n = 115). The rest of the sample was considered to be at low risk (based on the teacher's assessment of disruptiveness in kindergarten). The intervention took place when the boys were aged between 7 and 9 years old. GBTM was used to identify the trajectories for three different outcomes between 11 and 17 years old, that is physical aggression, vandalism and theft (based on self-reports). For each outcome, six trajectories were identified (i.e., Low 1, Low 2, Low rising, Low decline, Medium decline, and High rising). The authors examined the probabilities of following the trajectories of physical aggression, vandalism, and theft depending on whether the participants were in the intervention group, the control group or the low-risk group. Results suggested that the intervention changed the course of antisocial behavior throughout adolescence, especially physical aggression. Boys in the intervention group compared to those in the control group were more likely to follow the lowest-level trajectory of physical aggression and were less likely to follow high-level trajectories of physical aggression.

Using GMM, Muthén and colleagues (2002) examined the effects of a randomized intervention aimed at reducing aggressive behavior in children (119 boys were in the intervention group and 80 boys were in the control group). The intervention was applied during the first and second grade, and the children were followed from the first to seventh grade with respect to the course of aggressive behavior. The authors notably tested whether or not intervention effects differed across trajectories. They found that the estimated four-trajectory model fitted the data well (e.g., High, Medium, Low and Late-starters), and that the benefits from the intervention were more evident for boys who were initially more aggressive (i.e., the rate of change over time for the boys in the High trajectory was greater for the ones in the intervention group compared to their counterparts in the control group).

According to Muthén and colleagues (2002), trajectory-modeling techniques should not be used as a substitute for examining significant overall effects of an intervention (e.g., when the group that received the experimental treatment is compared with the control group on the outcomes of interest regardless of the trajectory membership). Reliance on trajectory modeling in the absence of overall effects of an intervention could result in spurious findings given the multiple comparisons issue (as more comparisons are performed, it becomes more likely that the treatment and control groups will appear to differ on at least one aspect by random chance alone). They recommended that trajectory modeling should be carried out by comparing the estimated trajectories with those from existing empirical data or theory. Still, the identification of different intervention effects for individuals following distinct trajectories can be clinically useful. For instance, this information could be used for designing intervention studies where specific interventions could be implemented to address the strengths and vulnerabilities of individuals following different trajectories based on longitudinal screening procedures (Muthén et al., 2002).

GCM to Test Predictors, Moderators and Mediators of Treatment Response Beauchaine and colleagues (2005) examined predictors, moderators, and mediators of treatment response among 514 3–8-year-old children treated for conduct problems. Children were randomly assigned to a waitlist condition or to one or more combinations of

parent, child, or teacher training. The study examined children's response to treatment using a latent GCM fit to three time points: (1) pre-treatment, (2) post-treatment, and (3) 1-year follow-up, where steeper negative slopes reflected greater improvement over time. The authors tested three questions: (1) For whom (or under what conditions) was treatment *in general* most effective (i.e., predictors of treatment response)? (2) For whom was a *particular* treatment most effective (i.e., moderators of treatment response)? (3) Why (by what mechanisms) did treatment work (i.e., mediators of treatment response)?

Regarding question 1 (predictors that described for whom treatment in general was most effective), the authors identified the following cases that were associated with better treatment response: older mother, father or mother with history of substance use, child with higher levels of anxiety/depression, and more treatment components. Regarding question 2 (moderators indicating for whom particular treatments were most effective), the authors identified the following cases for whom parent training was particularly effective: parents with lower marital satisfaction, depressed or single mother, father with history of substance use, and child with lower levels of anxiety/depression or from a lower SES family. The following cases for whom child training was particularly effective were: depressed mother, father with history of substance use, and a lower-SES family. Teacher training was particularly effective when the child was higher in attention problems. Regarding question 3 (mediators explaining treatment response), the authors identified reductions in the following parenting factors that explained why treatment worked: verbal criticism, harsh parenting, and ineffective parenting. In sum, GCMs can identify predictors, moderators, and mediators of treatment response.

Methodological Considerations

The next section provides an overview of relevant methodological considerations to take into account when applying trajectory modeling.

Measures of Behavior

Psychopathological disorders are often composed of different measured components. For instance, Conduct Disorder (American Psychiatric Association, 2013) is composed of criteria assessing aggression, destruction of property, deceitfulness or theft, and serious violations of rules. Aggregating these different types of behavior while modeling developmental trajectories may blur the analysis and mask etiological differences associated with the development of distinct types of behaviors. For instance, the emergence of physically aggressive behavior may be observed before theft and vandalism, which may emerge and show important growth during adolescence (Séguin, Sylvers, & Lilienfeld, 2007). Data suggest that physical aggression peaks during the second year of birth and then decreases (Tremblay, 2000). Let us consider Moffitt's taxonomy (Moffitt, 1993), which includes a lifecourse-persistent trajectory and an adolescence-limited trajectory. The identification of an adolescence-limited trajectory may depend on the items under investigation. For instance, if a scale of physically aggressive behavior is used, the likelihood of finding an adolescence-limited

trajectory may be low, whereas if a global measure of antisocial behavior (including physical aggression, but also other behaviors that may emerge later in development such as theft, vandalism, and drug use), the likelihood of finding an adolescence-limited trajectory may be high (Vitaro, Brendgen, & Barker, 2006). Developmental trajectories can vary across different types of behaviors (e.g., physical aggression, opposition, property violations and status violations; Bongers, Koot, van der Ende & Verhulst, 2004). This should be considered when modeling developmental trajectories of psychopathology.

Still, the best level of analysis is the most useful one for the researcher's goal. There are many ways to parse out behavior, which is heterogeneous by nature. For instance, externalizing problems can be split into aggression and non-aggression phenotypes. Aggression can further be reduced into physical, verbal, and relational. Physical aggression can be further reduced into proactive and reactive. Proactive physical aggression can be further reduced into targeting peers versus siblings. One could further reduce to the different brain mechanisms involved. At each level, the behaviors likely consist of heterogeneous mechanisms (and therefore combine behaviors with different developmental trajectories). One can always reduce to a lower level (more basic) unit, but the question is: what is the most useful level of analysis for the goal? This reflects the classic tension between lumpers and splitters. There are advantages of focusing on more specific behaviors, but there can also be benefits of accounting for the co-occurrence of different behaviors. For instance, there may be usefulness in examining the construct of general externalizing behavior because it is an efficient summary of many cases of psychopathology, the subdimensions tend to co-occur, and similar developmental processes appear to be involved with the different subdimensions (Olson, Bates, Sandy, & Lanthier, 2000). Alternatively, Insel (2014) has argued that instead of focusing on behavior to define syndromes (because behavior is imprecise and provides little information about the underlying mechanisms), we should focus on the developmental trajectories of underlying brain mechanisms. In sum, the researcher should consider his or her goal when choosing which levels of analysis to examine.

Number of Trajectories and Model Fit

An important aspect with regard to using GBTM or GMM concerns the determination of the number of trajectories that best represents or fits the data. As we previously mentioned, the use of GMM, compared to GBTM, could lead to a fewer number of trajectories, because adding random effects allows for more within-group variability in individual-level trajectory (Nagin & Odgers, 2010). Still, for both types of models, researchers should rely on criteria and theory to make their decision about the optimal number of groups.

Common criteria to assess the model fit include the Bayesian information criterion (BIC; Raftery, 1995), the Akaike information criterion (AIC; Akaike, 1974), the Lo–Mendell–Rubin likelihood ratio test (LMR-LRT; Lo, Mendell, & Rubin, 2001), and entropy (Muthén, 2004). The BIC and the AIC are indices that assess the model fit by balancing the goodness of fit to the sample data and the complexity of the model. They can be used to identify the best model from all the models considered (e.g., from a two-trajectory model to a six-trajectory model). That is, from the estimated models, the one with the lowest AIC (or the lowest BIC) would be selected as

best fitting the data. The LMR-LRT is an indicator to determine the ideal number of trajectories; a low *p*-value ($p < .05$) indicates that the k trajectory model is a better fit to the data compared to the k-1 trajectory model). Entropy is a measurement of classification accuracy, with values closer to 1 suggesting that the classes are well separated (range 0 to 1). Other tests of adequacy have been proposed such as verifying that the average of the posterior probabilities of trajectory membership for individuals assigned to each trajectory surpasses a minimum threshold (i.e., at least 0.70) (Nagin, 2005). Further information about the fit indices can be found in prior work (e.g., Nagin, 2005; Nylund, Asparouhov, & Muthén, 2007).

Model selection should also be based on the research questions/hypotheses under investigation, as well as extant relevant theories and prior empirical research. Although the identification of trajectories through GBTM or GMM can be clinically relevant, it is important to note that the identified trajectories only approximate the developmental course of subgroups in the population (Nagin & Odgers, 2010). Extant theories about the number and the shapes of trajectories for a given research question may not be available. In this case of a more exploratory framework, it becomes especially important to communicate and justify the decisions associated with the model selection. In addition, one strategy to verify if the identified trajectories appear meaningful is to test whether or not the trajectories can be differentiated from each other in terms of risk factors, response to treatment or outcomes. Replications of the findings using different datasets could also help to ascertain whether or not the trajectories identified represent meaningful subgroups of individuals in the population.

Although GCM does not involve the identification of distinct trajectory groups, it is still essential to establish the adequacy of fit for the growth models assessed (see Curran et al., 2010). The appropriate fit indices to be considered depend on the specific analytic strategy used to estimate the models (e.g., SEM or multilevel modeling). Examples include the RMSEA (root mean squared error of approximation) and the CFI (comparative fit index). Further information can be found in more technical references (e.g., Raudenbush & Bryk, 2002; Schreiber, Nora, Stage, Barlow, & King, 2006).

Data Sources

The types of data sources may impact the identification of the trajectories because of differences in the raters' assessments. For instance, teachers' ratings may be associated with higher independence in assessments (e.g., if there is a different teacher at each assessment) and are less likely to be influenced by the characteristics of any one rater (e.g., a parent's mental health problems; Côté, Zoccolillo, Tremblay, Nagin, & Vitaro, 2001). In addition, teachers and parents may underestimate certain behaviors, such as youth antisocial conduct (especially for teenagers), but self-reports can be affected by biases including problems recalling relevant information and social desirability. Self-reports may have the advantage of providing information that is unknown to adults or official services (e.g., police). For example, in a study on the links between self-reported and official offending behavior, GBTM was used to estimate trajectories of violent and nonviolent offending from self-reports collected between 11 and 17 years from a sample of 969 boys (Fontaine, Lacourse, Vitaro, & Tremblay, 2014). The researchers selected a five-trajectory model for violent offending (i.e., Chronic, 3%; Desisting, 12%; Delayed, 6%; Moderate, 47%; and Low, 32%) and a three-trajectory

model for nonviolent offending (Chronic, 8%; Moderate, 45%; and Low, 47%). Although they found that self-reports were associated with official records, especially for violent offending behavior, they could not perform meaningful trajectory analyses using the official offending data due to a low prevalence and frequencies of these behaviors, despite the use of an at-risk sample (i.e., boys from disadvantaged environments).

Indeed, the types of samples (e.g., population-based, at-risk, or clinical) should also be taken into account when performing the analyses and interpreting the findings, as it could influence the trajectories identified and the proportion of individuals assigned to each trajectory. For instance, the percentage of individuals following an early-onset/persistent trajectory of antisocial behavior has been found to be higher in at-risk or clinical samples compared to population-based samples (Fontaine, Carbonneau, Vitaro, Barker, & Tremblay, 2009).

Developmental Periods

To model trajectories (GCM, GBTM, and GMM), a minimum of three time points (usually as a function of age) is needed for proper estimation, but four or more time points are preferable to estimate more complex models, such as trajectories following quadratic or cubic trends (Curran & Muthén, 1999). This does not imply that all individuals need to have data on all time points (see the section on missing data management, p. 22). However, several time points across different developmental periods may be necessary to address certain research questions. For instance, if a researcher aims to test a developmental theoretical model, such as Moffitt's taxonomy (i.e., to test whether or not lifecourse-persistent and adolescence-limited trajectories can be identified; Moffitt, 1993), data collected at different time points from childhood to adulthood would be necessary. A number of published studies on the developmental trajectories of antisocial behavior were based on data collected at the end of childhood or during adolescence (i.e., where a proportion of individuals already manifested behavioral problems) and did not include data in adulthood (e.g., Broidy et al., 2003; Fontaine et al., 2014). Thus, the empirical validation of such a model on the development of individuals' antisocial behavior may be limited due to restrictive ranges of the ages considered.

The number and the shape of each trajectory are not fixed certainties. The number of time points or periods for which the individuals are monitored can have an impact on the number and shapes of the trajectories, although this remains an understudied topic (Nagin, 2005). For instance, Eggleston, Laub, and Sampson (2004) tested whether or not length of follow-up affected trajectory number, shape and group membership. Using a sample of 500 delinquent boys and their official crime counts from ages 7 to 70, their findings suggested that the length of follow-up (7 to 25 vs. 7 to 32 vs. 7 to 45 vs. 7 to 70) influenced trajectory number, shapes, and group membership, but that groups were differentially affected. The high rate-chronic group, for example, was not identified when the analysis was restricted to the shortest age range (i.e., 7 to 25). Eggleston and colleagues (2004) also reported that variables such as incarceration and mortality could affect the estimation of offending patterns over time. For instance, individuals who are incarcerated may be classified in a desisting offending trajectory because they would not show up in official records. Similarly, trajectory analyses on depressive symptoms may classify participants who passed away as desisters.

Therefore, it appears important to take into account the length of follow-up, the type of variables under investigation (e.g., offending behavior) and to consider that the trajectories are not immutable. In addition, it appears that more data allow for more refined statistical inferences (Nagin, 2005). For instance, in clinical research, failing to identify trajectories because of too few time points could lead to errors in intervention effectiveness given that heterogeneity may wash out effects for some or identify positive effects for others. Further research is needed to test further the influence of temporal length of studies on the trajectory estimation, as well as other design features such as spacing of temporal retest intervals and sample size.

Sample Size

The sample size, that is the number of individuals included in the analysis, can also have an impact on the estimation of the trajectories. It seems that a clear consensus has not been established about the minimum number of individuals needed to perform trajectory analysis. In GCM, sample sizes approaching at least 100 are often preferred (Curran et al., 2010). However, we should expect that large samples of individuals followed throughout several years (and developmental periods), allow for greater capacity to identify meaningful trajectories (Hertzog, Lindenberger, Ghisletta, & von Oertzen, 2006; Nylund et al., 2007). Indeed, one may question the meaningfulness of having one or two trajectories with a very small number of participants (e.g., a high chronic trajectory including five participants when using GBTM or GMM). Using GBTM, D'Unger, Land, McCall, and Nagin (1998) tested whether or not the number of trajectories in the selected model was influenced by the number of individuals included in the analysis (i.e., samples of 500, 1,000 and 2,000 individuals). They found that their selected model (i.e., the model with five trajectories of offenders) was robust to sample size. Sampson, Laub, and Eggleston (2004) varied their sample size from 25 men to almost 500 to model trajectories of offending behavior. They reported that the number of trajectories varied from five to eight, with the eight-trajectory model derived only for sample sizes of 200 and higher. Sampson and colleagues (2004) suggested that the sample size problem is relatively modest. Theories and previous empirical research could be used to inform design decisions when planning a longitudinal study, including the required sample size that may best allowed to test the hypotheses under investigation, such as the number and shapes of trajectories.

Missing Data

When using longitudinal data, researchers and clinicians are often confronted with missing information and attrition. Several reasons can lead to missing data across assessments, including participants' refusal to respond to specific questions of the assessment and participants dropping out of clinical experimentations and being lost within longitudinal studies (because of refusal to continue to participate, disappearance, or death). The impact of missing data when performing trajectory analysis can be particularly problematic if missingness is related to the variables of interest. For instance, in research on antisocial behavior, individuals with the highest levels of adjustment problems may also tend to have the highest levels of missing data

(e.g., adolescents with antisocial behavior may be unwilling to continue participation in a study into which their parents initially enrolled them, and initially cooperative parents may drop out at later ages of their children even if their children remain in the study; Asendorpf, van de Schoot, Denissen, & Hutteman, 2014). Biases can be introduced if individuals with incomplete data are excluded from the analyses, especially if they differ on key characteristics from those who are included.

One important step when applying trajectory modeling is to document and understand both the type and the quantity of missing data across assessments (Nagin & Odgers, 2010). Missing data can be classified into different categories. They can be characterized as missing completely at random (data are truly missing at random), as missing at random (data are missing as a function of measured characteristics, such as sex) or as missing not at random (data are missing as a function of unmeasured characteristics such as the very value that is missing) (Curran et al., 2010). The statistical importance and complexity of the missing data management will depend on the type of missing data (Little & Rubins, 1987; Schafer & Graham, 2002). A detailed discussion of the different methods of missing data management would go beyond the scope of the present chapter. However, it should be noted that missing data issues can be addressed using statistical techniques either outside of the trajectory modeling framework (e.g., multiple imputation) or when performing the trajectory analysis (e.g., when using the full-information maximum likelihood estimation) (for more information, see Enders, 2001; Nagin & Odgers, 2010).

Conclusions

In this chapter, we presented an overview of three trajectory modeling approaches, namely GCM, GBTM, and GMM. These approaches share a common analytical goal: they allow the modeling of differences or variability across members of a population in their developmental trajectories (Nagin & Odgers, 2010). We also presented examples of applications of these trajectory modeling approaches in the field of developmental psychopathology and clinical psychology, such as studies that focused on the development of externalizing and internalizing problems, on the validation of theoretical models and on the examination of intervention effects. Finally, we discussed methodological considerations when applying these models and interpreting the findings, such as issues related to the measurement of behavior, the determination of the number of trajectories that best represents or fits the data, the potential impact of data sources, the influence of the developmental periods considered, the requirement in terms of sample size and the management of missing data.

Numerous research questions about the developmental trajectories of psychopathology can be addressed using these approaches. Researchers need to select the approach that will allow them to address appropriately their hypotheses given the data to be analyzed. For instance, if they expect that all individuals in their sample come from the same population and can be described by the same parameters of change, then GCM could be appropriate. Alternatively, if they expect that some individuals come from distinct subpopulations, which can be captured by different subgroups, GBTM (with the assumption that the subgroups are homogeneous regarding change) and GMM (if heterogeneity within each subgroup is considered) could be more

appropriate. Researchers should clearly articulate their rationale for using the approach they have selected and they should also justify their final model (e.g., choice of the number of trajectories) based on a combination of formal statistical criteria as well as extant empirical data and theory.

Several software packages for fitting these models are available, including SAS (PROC TRAJ, for GBTM; Nagin, 2005), Mplus (e.g., GMM and GCM; Muthén & Muthén, 2012), and R (SEM, Boker et al., 2011; HLM, Pinheiro, Bates, DebRoy, Sarkar, & the R Core team, 2009). Tutorials and workshops that focus on the theory and technical applications of these models are also available and could be useful to get started or to acquire more advanced knowledge. We presented an overview of these approaches and applications. Several recent published articles, chapters, and textbooks cover more comprehensive and technical aspects of these approaches and their applications (e.g., Hox, 2010; Little, 2013; McArdle & Nesselroade, 2014; Muthén, 2004; Nagin, 2005; Nylund et al., 2007; Raudenbush & Bryk, 2002; Shiyko, Ram, & Grimm, 2012; Singer & Willett, 2003). Statistical models are constantly evolving; collaborations between researchers, clinicians and developers of these models should lead to enhanced tools for testing hypotheses in the domains of developmental psychopathology and clinical psychology.

Acknowledgments

We thank Mr. Alain Girard, statistician, for the helpful comments on this chapter. Dr. Nathalie Fontaine is a Research Scholar, Junior 1, Fonds de recherche du Québec–Santé.

References

Achenbach, T. M. (2009). *Achenbach System of Empirically Based Assessment (ASEBA): Development, findings, theory, and applications.* University of Vermont, Research Center of Children, Youth & Families.

Akaike, H. (1974). New look at statistical-model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th Ed.). Washington, DC: Author.

Asendorpf, J. B., van de Schoot, R., Denissen, J. J. A., & Hutteman, R. (2014). Reducing bias due to systematic attrition in longitudinal studies: The benefits of multiple imputation. *International Journal of Behavioral Development*, *38*, 453–460. doi:10.1177/0165025414542713

Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics*, *28*, 135–167. doi:10.3102/10769986028002135

Bauer, D. J., & Reyes, H. L. M. (2010). Modeling variability in individual development: Differences of degree of kind? *Child Development Perspectives*, *4*, 114–122. doi:10.1111/j.1750-8606.2010.00129.x

Beauchaine, T. P., Webster-Stratton, C., & Reid, M. J. (2005). Mediators, moderators, and predictors of 1-year outcomes among children treated for early-onset conduct problems: A latent growth curve analysis. *Journal of Consulting and Clinical Psychology*, *73*, 371–388. doi:10.1037/0022-006X.73.3.371

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., Spies, J., Estabrook, R., Kenny, S., Bates, T., Mehta, P., & Fox, J. (2011). OpenMx: An open source extended structural equation modeling framework. *Psychometrika*, *76*, 306–317. doi:10.1007/s11336-010-9200-6

Bongers, I. L., Koot, H. M., van der Ende, J., & Verhulst, F. C. (2003). The normative development of child and adolescent problem behavior. *Journal of Abnormal Psychology*, *112*, 179–192. doi:10.1037/0021-843x.112.2.179

Bongers, I. L., Koot, H. M., van der Ende, J., & Verhulst, F. C. (2004). Developmental trajectories of externalizing behaviors in childhood and adolescence. *Child Development*, *75*, 1523–1537. doi:10.1111/j.1467-8624.2004.00755.x

Broidy, L. M., Nagin, D. S., Tremblay, R.E., Bates, J. E., Brame, B., Dodge, K. A., Fergusson, D., Horwood, J. L., Loeber, R., Laird, R., Lynam, D. R., Moffitt, T. E., Pettit, G. S., & Vitaro, F. (2003). Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study. *Developmental Psychology*, *39*, 222–245. doi:10.1037/0012-1649.39.2.222

Cicchetti, D. (1989). Developmental psychopathology: Some thoughts on its evolution. *Development and Psychopathology*, *1*, 1–4. doi:10.1017/S0954579400000201

Cicchetti, D., & Toth, S. L. (1992). The role of developmental theory in prevention and intervention. *Development and Psychopathology*, *4*, 489–493. doi:10.1017/S0954579400004831

Coghill, D., & Sonuga-Barke, E. J. S. (2012). Annual research review: Categories versus dimensions in the classification and conceptualisation of child and adolescent mental disorders: implications of recent empirical study. *Journal of Child Psychology and Psychiatry*, *53*, 469–489. doi:10.1111/j.1469-7610.2011.02511.x

Côté, S. M., Boivin, M., Liu, X., Nagin, D. S., Zoccolillo, M., & Tremblay, R. E. (2009). Depression and anxiety symptoms: Onset, develomental course and risk factors during early childhood. *Journal of Child Psychology and Psychiatry*, *50*, 1201–1208. doi:10.1111/j.1469-7610.2009.02099.x

Côté, S., Zoccolillo, M., Tremblay, R. E., Nagin, D., & Vitaro, F. (2001). Predicting girls' conduct disorder in adolescence from childhood trajectories of disruptive behaviors. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*, 678–684. doi: 10.1097/00004583-200106000-00013

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research*, *38*, 529–569. doi:10.1207/s15327906mbr3804_5

Curran, P. J., & Muthén, B. O. (1999). The application of latent curve analysis to testing developmental theories in intervention research. *American Journal of Community Psychology*, *27*, 567–595. doi:10.1023/A:1022137429115

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development*, *11*, 121–136. doi:10.1080/15248371003699969

D'Unger, A. V., Land, K. C., McCall, P. L., & Nagin, D. S. (1998). How many latent classes of delinquent/criminal careers? Results from mixed Poisson regression analyses. *American Journal of Sociology*, *103*, 1593–1630.

Eggleston, E. P., Laub, J. H., & Sampson, R. J. (2004). Methodological sensitivities to latent class analysis of long-term criminal trajectories. *Journal of Quantitative Criminology*, *20*, 1–26. doi:10.1023/B:JOQC.0000016696.02763.ce

Enders, C. K. (2001). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling*, *8*, 128–141. doi:10.1207/S15328007SEM0801_7

Fairchild, G., van Goozen, S. H. M., Calder, A. J., & Goodyer, I. M. (2013). Research review: Evaluating and reformulating the developmental taxonomic theory of antisocial behaviour. *Journal of Child Psychology and Psychiatry*, *54*, 924–940. doi:10.1111/jcpp.12102

Fontaine, N., Carbonneau, R., Barker, E. D., Vitaro, F., Hébert, M., Côté, S. M., Nagin, D. S., Zoccolillo, M., & Tremblay, R. E. (2008). Girls' hyperactivity and physical aggression during

childhood and adjustment problems in early adulthood: A 15-year longitudinal study. *Archives of General Psychiatry*, *65*, 320–328. doi:10.1001/archgenpsychiatry.2007.41

Fontaine, N., Carbonneau, R., Vitaro, F., Barker, E. D., & Tremblay, R. E. (2009). Research review: A critical review of studies on the developmental trajectories of antisocial behavior in females. *Journal of Child Psychology and Psychiatry*, *50*, 363–385. doi:10.1111/j.1469-7610.2008.01949.x

Fontaine, N. M. G., Lacourse, E., Vitaro, F., & Tremblay, R. E. (2014). Links between trajectories of self-reported violent and nonviolent offending and official offending during adolescence and adulthood. *Criminal Behaviour and Mental Health*, *24*, 277–290. doi:10.1002/cbm.1935

Fontaine, N. M. G., Rijsdijk, F. V., McCrory, E. J. P., & Viding, E. (2010). Etiology of different developmental trajectories of callous–unemotional traits. *Journal of the American Academy of Child and Adolescent Psychiatry*, *49*, 656–664. doi:10.1016/j.jaac.2010.03.014

Grimm, K. J., Ram, N., & Hamagami, F. (2011). Nonlinear growth curves in developmental research. *Child Development*, *82*, 1357–1371. doi:10.1111/j.1467-8624.2011.01630.x

Gueorguieva, R., & Krystal, J. H. (2004). Move over ANOVA: Progress in analyzing repeated-measures data and its reflection in papers published in the archives of general psychiatry. *Archives of General Psychiatry*, *61*, 310–317. doi:10.1001/archpsyc.61.3.310

Hertzog, C., Lindenberger, U., Ghisletta, P., & von Oertzen, T. (2006). On the power of multivariate latent growth curve models to detect correlated change. *Psychological Methods*, *11*, 244–252. doi:10.1037/1082-989X.11.3.244

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd Ed.). New York, NY: Routledge.

Hu, M.-C., Muthén, B., Schaffran, C., Griesler, P. C., & Kandel, D. B. (2008). Developmental trajectories of criteria of nicotine dependence in adolescence. *Drug and Alcohol Dependence*, *98*, 94–104. doi:10.1016/j.drugalcdep.2008.04.017

Insel, T. R. (2014). Mental disorders in childhood: Shifting the focus from behavioral symptoms to neurodevelopmental trajectories. *JAMA: Journal of the American Medical Association*, *311*, 1727–1728. doi:10.1001/jama.2014.1193

Keiley, M. K., Bates, J. E., Dodge, K. A., & Pettit, G. S. (2000). A cross-domain growth analysis: externalizing and internalizing behaviors during 8 years of childhood. *Journal of Abnormal Child Psychology*, *28*, 161–179. doi:10.1023/A:1005122814723

Krueger, R. F., Markon, K. E., Patrick, C. J., & Iacono, W. G. (2005). Externalizing psychopathology in adulthood: A dimensional-spectrum conceptualization and its implications for DSM-V. *Journal of Abnormal Psychology*, *114*, 537–550.

Lacourse, E., Côté, S., Nagin, D. S., Vitaro, F., Brendgen, M., & Tremblay, R. E. (2002). A longitudinal-experimental approach to testing theories of antisocial behavior development. *Development and Psychopathology*, *14*, 909–924. doi:10.1017/S0954579402004121

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: The Guilford Press.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*, 767–778. doi:10.1093/biomet/88.3.767

Markon, K. E., & Krueger, R. F. (2005). Categorical and continuous models of liability to externalizing disorders: A direct comparison in NESARC. *Archives of General Psychiatry*, *62*, 1352–1359. doi:10.1001/archpsyc.62.12.1352

McArdle, J. J., & Nesselroade, J. R. (2014). *Longitudinal data analysis using structural equation models*. Washington, DC: American Psychological Association.

Miller, J., Vaillancourt, T. & Boyle, M. (2009). Examining the heterotypic continuity of aggression using teacher reports: Results from a national Canadian study. *Social Development*, *18*, 164–180. doi:10.1111/j.1467-9507.2008.00480.x

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review*, *100*, 674–701. doi:10.1.1.314.783

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 345–368). Thousand Oaks, CA: Sage.

Muthén, B., Brown, C. H., Masyn, K., Jo, B., Khoo, S. T., Yang, C. C., Wang, C. P., Kellam, S. G., Carlin, J. B., & Liao, J. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics*, *3*, 459–475. doi:10.1093/biostatistics/3.4.459

Muthén, B. & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, *55*, 463–469. doi:10.1111/j.0006-341X.1999.00463.x

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th Ed.). Los Angeles, CA: Muthén & Muthén.

Nærde, A., Ogden, T., Janson, H., & Zachrisson, H. D. (2014). Normative development of physical aggression from 8 to 26 months. *Developmental Psychology*, *50*, 1710–1720. doi:10.1037/a0036324

Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods*, *4*, 139–157. doi:10.1037/1082-989X.4.2.139

Nagin, D. S. (2005). *Group-based modeling of development*. Cambridge, MA: Harvard University Press.

Nagin, D. S., & Odgers, C. L. (2010). Group-based trajectory modeling in clinical research. *Annual Review of Clinical Psychology*, *6*, 109–138. doi:10.1146/annurev.clinpsy.121208.131413

Nagin, D. & Tremblay, R. E. (1999). Trajectories of boys' physical aggression, opposition, and hyperactivity on the path to physically violent and nonviolent juvenile delinquency. *Child Development*, *70*, 1181–1196. doi:10.1111/1467-8624.00086

Nagin, D. S. & Tremblay, R. E. (2001). Analyzing developmental trajectories of distinct but related behaviors: A group-based method. *Psychological Methods*, *6*, 18–34. doi:10.1037/1082-989X.6.1.18

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling. A Monte Carlo simulation study. *Structural Equation Modeling*, *14*, 535–569. doi:10.1080/10705510701575396

Odgers, C. L., Moffitt, T. E., Broadbent, J. M., Dickson, N., Hancox, R. J., Harrington, H., Poulton, R., Sears, M. R., Thomson, W. M., & Caspi, A. (2008). Female and male antisocial trajectories: From childhood origins to adult outcomes. *Development and Psychopathology*, *20*, 673–716. doi:10.1017/S0954579408000333

Olson, S. L., Bates, J. E., Sandy, J. M., & Lanthier, R. (2000). Early developmental precursors of externalizing behavior in middle childhood and adolescence. *Journal of Abnormal Child Psychology*, *28*, 119–133. doi:10.1023/A:1005166629744

Owens, E. B., & Shaw, D. S. (2003). Predicting growth curves of externalizing behavior across the preschool years. *Journal of Abnormal Child Psychology*, *31*, 575–590. doi:10.1023/a:1026254005632

Petersen, I. T., Bates, J. E., Dodge, K. A., Lansford, J. E., & Pettit, G. S. (2015). Describing and predicting developmental profiles of externalizing problems from childhood to adulthood. *Development and Psychopathology*, *27*, 791–818. doi: 10.1017/S0954579414000789

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & the R Core team. (2009). nlme: Linear and nonlinear mixed effects models. *R package version 3*.1–93. Retrieved December 9, 2016 from http://cran.r-project.org/web/packages/nlme/index.html.

Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological methodology* (Vol. *25*, pp. 111–163). Washington, DC: American Sociological Association.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis method* (2nd Ed.). Thousand Oaks, CA: Sage Publications.

Rutter, M. (1990). Psychosocial resilience and the protective mechanisms. In J. Rolf, A. S. Masten, D. Cicchetti, K. H. Nuechterlein, & S. Weintraub (Eds.), *Risk and protective factors in the development of psychopathology* (pp. 181–214). Cambridge: Cambridge University Press.

Sampson, R. J., Laub, J. H., & Eggleston, E. P. (2004). On the robustness and validity of groups. *Journal of Quantitative Criminology*, *20*, 37–42. doi:10.1023/B:JOQC.0000016698.36239.91

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147–177. doi:10.1037/1082-989x.7.2.147

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *Journal of Educational Research*, *99*, 323–337. doi:10.3200/JOER.99.6.323-338

Séguin, J. R., Sylvers, P., & Lilienfeld, S. O. (2007). The neuropsychology of violence. In D. J. Flannery, A. T. Vazsonyi, & I. D. Waldman (Eds.), *The Cambridge handbook of violent behavior and aggression* (pp. 187–214). Cambridge: Cambridge University Press.

Shiyko, M., Ram, N., & Grimm, K. J. (2012). An overview of growth mixture modelling: A sample nonlinear application in OpenMx. In R. H. Hoyle (Ed.), *Handbook of structural equation modelling* (pp. 532–546). New York: The Guilford Press.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Tremblay, R. E. (2000). The development of aggressive behaviour during childhood: What have we learned in the past century? *International Journal of Behavioral Development*, *24*, 129–141. doi:10.1080/016502500383232

Viding, E., Fontaine, N. M. G., & McCrory, E. J. (2012). Antisocial behaviour in children with and without callous–unemotional traits. *Journal of the Royal Society of Medicine*, *105*, 195–200. doi:10.1258/jrsm.2011.110223

Vitaro, F., Brendgen, M., & Barker, E. D. (2006). Subtypes of aggressive behaviors: A developmental perspective. *International Journal of Behavioral Development*, *30*, 12–19. doi:10.1177/0165025406059968

Walters, G. D. (2011). The latent structure of life-course-persistent antisocial behavior: Is Moffitt's developmental taxonomy a true taxonomy? *Journal of Consulting and Clinical Psychology*, *79*, 96–105. doi:10.1037/a0021519

Walters, G. D. (2012). Developmental trajectories of delinquent behavior: One pattern or several? *Criminal Justice and Behavior*, *39*, 1192–1203. doi:10.1177/0093854812443645

Walters, G. D., & Ruscio, J. (2013). Trajectories of youthful antisocial behavior: Categories or continua? *Journal of Abnormal Child Psychology*, *41*, 653–666. doi:10.1007/s10802-012-9700-1

Walton, K., Ormel, J., & Krueger, R. (2011). The dimensional nature of externalizing behaviors in adolescence: Evidence from a direct comparison of categorical, dimensional, and hybrid models. *Journal of Abnormal Child Psychology*, *39*, 553–561. doi:10.1007/s10802-010-9478-y

2

Family-based Quasi-experimental Designs for Studying Environmental Risk Factors

Henrik Larsson and Brian M. D'Onofrio

Recent research that has found correlations between early environmental risk factors and psychopathology (Bale et al., 2010; Beydoun & Saftlas, 2008; Buss, Davis, Muftuler, Head, & Sandman, 2010; A. C. Huizink, Robles de Medina, Mulder, Visser, & Buitelaar, 2003; O'Connor, Heron, Golding, Beveridge, & Glover, 2002; O'Connor, Heron, Golding, Glover, & ALSPAC study team, 2003) has led to great excitement about the developmental origins of disease hypothesis, which emphasizes the significant role of early stressors for later functioning (Barker, 1998; Gluckman & Hanson, 2007; Gluckman, Hanson, Cooper, & Thornburg, 2008). Yet, many researchers have expressed concern of such strong causal inferences because most of the studies have not been able to rigorously test causal inferences by ruling out plausible alternative explanations for the associations between the risk factors and outcomes (Kramer, 2000; O'Connor, 2003; Thapar & Rutter, 2009).

Because we cannot randomly assign individuals to most of the early risk factors for psychopathology, researchers must rely on alternatives to randomized controlled trials for ruling out plausible competing explanations for the associations (Kraemer et al., 1997; Shadish, Cook, & Campbell, 2002). Given that it is impossible to identify and accurately measure every important confounding factor, there is a growing push in medicine (Academy of Medical Sciences Working Group, 2007; British Academy of Science Working Group, 2010), psychiatry (Kendler, 2005; Lahey, D'Onofrio, & Waldman, 2009), statistics (Shadish, 2010), sociology (Freese, 2008), psychology (Rutter, Pickles, Murray, & Eaves, 2001; Shadish et al., 2002), family studies (D'Onofrio & Lahey, 2010), epidemiology (Donovan & Susser, 2011; Gilman & Loucks, 2012), basic sciences (Smith, 2008), and economics (Duncan, 2012), to use quasi-experimental designs, approaches that rely on design features to help rule out alternative hypotheses, such as unmeasured environmental confounding factors (Rutter, 2000), unmeasured genetic confounding factors (due to gene–environment correlations) (Kendler & Baker, 2007; Plomin & Bergeman, 1991), and reciprocal influences/child effects (Bell & Harper, 1997). And, reviews have specifically called for more quasi-experimental studies of early environmental risk factors for

The Wiley Handbook of Developmental Psychopathology, First Edition. Edited by Luna C. Centifanti and David M. Williams.

psychopathology because of the limitations of existing studies (Duncan, 2012; O'Connor, 2003; Pluess & Belsky, 2011; Rutter, 2005; Rutter, 2007; Shonkoff, Boyce, & McEwen, 2009).

In this context, this chapter stresses the advantages of using family-based, quasi-experimental research designs when studying early environmental risk factors for psychopathology because the approaches allow strong tests of causal inferences that are critical for advancing the understanding of the etiology of psychopathology. We use recent quasi-experimental studies on the association between maternal smoking during pregnancy (SDP) and Attention Deficit Hyperactivity Disorder (ADHD) as an exemplar.

ADHD is a stable (Faraone, Biederman, & Mick, 2006) neurodevelopmental disorder with an early onset that predicts important social outcomes later in life (Biederman et al., 2010). Even though the secular trend in ADHD (Singh, 2008) has been debated, current estimates of the prevalence of ADHD during childhood (≈5%) (Polanczyk, de Lima, Horta, Biederman, & Rohde, 2007) further underscore their large societal impact. Understanding the etiology of this disorder, therefore, is of critical importance. In addition to a strong genetic component, quantitative genetic studies highlight that environmental factors are important (Faraone et al., 2005; Larsson, Anckarsater, Rastam, Chang, & Lichtenstein, 2012). The early onset of ADHD and early emerging structural and functional brain abnormalities suggest a prenatal and early postnatal origin of ADHD (Swanson et al., 2007). Recent research has indeed identified several early risk factors for ADHD (Halperin & Schulz, 2006; Thapar, Cooper, Eyre, & Langley, 2013), consistent with the developmental origins of health and disease hypothesis that emphasizes the significant role of early stressors for later functioning (Bale et al., 2010; Szatmari, 2011). Among these putative environmental risk factors maternal smoking during pregnancy (SDP) is probably the most commonly cited risk factor for ADHD (Froehlich et al., 2009; Galera et al., 2011; Mick, Biederman, Faraone, Sayer, & Kleinman, 2002; Milberger, Biederman, Faraone, & Jones, 1998; C. Obel et al., 2009; Rodriguez & Bohlin, 2005b).

SDP is associated with various offspring problems, including pregnancy-related and perinatal outcomes (Cnattingius, 2004) and later developmental outcomes, such as cognitive deficits (Lambe, Hultman, Torrang, MacCabe, & Cnattingius; Lundberg et al., 2010), obesity (Iliadou et al., 2010) and social and behavioural difficulties (Glantz & Campbell Champers, 2006; Wakschlag, Pickett, Cook, Benowitz, & Leventhal, 2002). Several reviews have concluded that fetal exposure to maternal SDP causes these problems (Cnattingius, 2004; M. Y. Glantz & Campbell Champers, 2006; Slotkin, 2013; Wakschlag et al., 2002) because the statistical associations (a) have been replicated in numerous studies, (b) are independent of the effects of measured covariates that are used to account for confounding, and (c) and consistent with basic neuroscience research on SDP in animals (Ernst, 2001; Shea & Steiner, 2008). Based in this body of literature researchers have hypothesized mediating mechanisms through which maternal SDP influences offspring psychopathology (Cornelius, Ryan, Day, Godschmidt, & Willford, 2001). For instance, researchers suggest that maternal SDP alters fetal development of brain systems related to stress reactivity (Koob, 1999), reward sensitivity (Ferriero & Dempsey, 1999; M. D. Glantz & Chambers, 2006), and decision-making (Lotfipour et al., 2009), which could subsequently influence offspring psychopathology, including ADHD. Yet, SDP is correlated with numerous risk factors, such as maternal and paternal intellectual abilities, psychopathology

(including ADHD), substance use problems, and socioeconomic status that also predict offspring ADHD (Huizink & Mulder, 2006). As such, an association between SDP and offspring psychopathology could be explained by unmeasured background factors rather by casual effects of SDP. Quantitative genetic research has also found that maternal SDP is influenced by genetic factors (Agrawal et al., 2008; D'Onofrio et al., 2003; Ellingston, Rickert, Lichtenstein, Långström, & D'Onofrio, 2012). The genetic factors that influence SDP are correlated with both nicotine dependence problems (Agrawal et al., 2008) and criminality (Ellingston et al., 2012). The fact that genetic factors influence SDP, therefore, raises the possibility that genetic confounds also could account for the association between SDP and offspring psychopathology, including ADHD (Rutter et al., 2001; Scarr & McCartney, 1983). Given these alternative explanations and the fact that few human studies are able to account for all environmental and genetic confounds, several researchers have explicitly cautioned against drawing strong causal inferences related to the consequences of maternal SDP (D'Onofrio et al., 2003; Fergusson, 1999; Maughan, Taylor, Caspi, & Moffitt, 2004; Silberg et al., 2003; Wakschlag et al., 2002).

Confounding in Observational Studies

Most observational studies use unrelated individuals (i.e., one person per family). This means that environmental risk factors can be confounded with genetic factors and environmental influences that are shared within the family. Confounds between genetic variants and environments, which are pervasive (Jaffee & Price, 2012; Kendler & Baker, 2007; Plomin & Bergeman, 1991), can arise systematically through two basic types of gene–environment correlation (rGE) (Eaves, Last, Martin, & Jinks, 1977; Plomin, DeFries, & Loehlin, 1977; Scarr & McCartney, 1983):

1 *Passive rGE.* This occurs because parents provide both their children's segregating genes and home environments. When an allele (version) of a genetic variant in parents is associated with their childrearing, the same allele in the offspring is passively correlated with their experienced childrearing.
2 *Active and evocative rGE.* Genes and environments also become correlated when the genetically influenced behavior and characteristics of individuals actively select them into, or evoke changes in, their environments.

Family-Based Quasi-Experimental Designs for Environmental Risk Factors

Although there are numerous designs that can strengthen causal inferences by ruling out different forms of confounding (Rutter et al., 2001), including confounding due to genetic factors, this chapter focuses on family-based quasi-experimental designs. In general, quasi-experiments are based on studies in which individuals are not randomly assigned to conditions but that use *design features* to rule out confounding factors instead of solely using statistical controls for measured covariates (Shadish et al., 2002). Family-based quasi-experimental designs in particular are based on comparing family members who vary in their exposures and outcomes (D'Onofrio, Lahey,

Turkheimer, & Lichtenstein, 2013). The approaches provide rigorous tests of causal inferences by delineating among different plausible explanations for statistical associations between risk factors and outcomes. It is important to note that each approach has assumptions and limitations that influence their ability to make strong causal inferences (internal validity) and generalize to other samples (external validity). As such, using multiple family-based quasi-experimental designs can help researchers identify the mechanisms underlying the statistical associations between risk factors and outcomes. We briefly review several designs, including sibling-comparison, offspring of siblings/twins, adoptions, and the *in vitro* fertilization designs here; in-depth reviews of the designs and others are available elsewhere (D'Onofrio et al., 2013; Eaves, Last, Young, & Martin, 1978; Lawlor & Mishra, 2009; Rutter et al., 2001). We will also briefly review the possibilities when researchers includes *multiple* family relationships that differ in their environmental exposure and genetic risk in the same study because quantitative genetic models can be used to further test competing causal hypotheses (D'Onofrio et al., 2003; Heath et al., 1993; Schermerhorn et al., 2011).

Sibling-comparison Designs

Instead of comparing unrelated individuals who vary in their exposure to a risk (e.g., SDP), which is the typical design in developmental psychopathology studies, sibling-comparison studies explore differences among siblings who were differentially exposed (e.g., one sibling was exposed to SDP but his/her sibling was not). These comparisons help account for certain types of confounding because siblings share familial factors that could confound the associations between the risk and outcome. For example, the comparisons of siblings raised in the same family account for all environmental confounds that are shared by siblings, such as early life shared environmental factors. Furthermore, sibling-comparisons help rule out some forms of genetic confounding (when genetic factors influence exposure to the risk and outcome) (Donovan & Susser, 2011; Lahey & D'Onofrio, 2010; Rutter, 2007). The process of meiosis, the type of cell division that produces eggs and sperm, randomly distributes alleles from parents to each of their offspring. As such, this random process rules out the *systematic* genetic confounding due to passive rGE. For example, the comparison of full-siblings (siblings who share 50% of their genes), rules out the possibility that genetic factors passively passed down from both parent could account for the association between the exposure and the outcome. The design does not rule out genetic confounding arising from active or evocative rGE, which occurs when a child's genotype influences the exposure to the risk (and can vary within siblings). To completely rule out all genetic factors researchers can compare identical twins who share 100% of their genetic sequences (McGue, Osler, & Christensen, 2010). It is important to note that the design cannot be used to examine exposures that are shared by twins, such as SDP (Purcell & Koenen, 2005; Turkheimer, D'Onofrio, Maes, & Eaves, 2005). However, if the probability of exposure cannot be influenced by the genetic characteristics of the offspring (e.g., the risk occurred before conception, such as maternal or paternal age at childbearing), the combination of the temporal ordering and sibling-comparisons can rule out all genetic confounding.

Like all research designs, sibling-comparisons have a number of assumptions and limitations. First, the designs cannot rule out environmental confounds that vary within siblings. Second, the comparison of full-siblings cannot always identify which

confounding factors responsible for associations between risk factors and outcomes (Donovan & Susser, 2011; Lahey & D'Onofrio, 2010). It is important to note however, that researchers can combine different sibling-comparison designs (e.g., full- *and* half-siblings) (Tierney, Merikangas, & Risch, 1994) or fraternal *and* identical twins (Neale & Cardon, 1992)) and quantitative genetic models to explore the degree to which an association is due to genetic and shared environmental confounds or environments that make siblings unique. If the association was due to environments that make siblings unique the results would be consistent with a causal inference (Turkheimer & Harden, 2014).

Third, researchers have to consider whether the results from the comparison of differentially exposed siblings generalize to other populations (Frisell, Oberg, Kuja-Halkola, & Sjolander, 2012; Lahey & D'Onofrio, 2010). Fourth, sibling-comparisons rely on the assumptions that there are no carry-over effects, which occurs when the exposure of one sibling influences the outcome of another (Donovan & Susser, 2011; Frisell et al., 2012; Lahey & D'Onofrio, 2010; Rutter, 2007; Susser, Eide, & Begg, 2010). Fortunately, researchers can test these assumptions by using different quasi-experimental approaches. For instance, researchers can conduct bidirectional case-cross studies, which explore differential exposure within sibling pairs based on their birth order (i.e., when the mother smoked more during the first versus the second pregnancy), to test the assumption of no carry-over effects (Meyer, Williams, Hernandez-Diaz, & Cnattingius, 2004).

Offspring of Siblings/Twins

Instead of comparing differentially exposed siblings, the offspring of siblings/twins compare differentially exposed cousins to help account for unmeasured genetic and environmental selection factors. The comparisons account for environmental effects that make cousins similar. Cousin-comparisons also can help account for genetic selection, although cousins share much less of their genetic makeup than siblings. For example, offspring of half-siblings (6.25% of genetic variants), offspring of full-siblings and fraternal twins (12.5%), and offspring of identical twins (25%) share less genetic makeup than full siblings. Interestingly, the offspring of identical twins design share the same amount of genetic makeup as half-siblings, so that socially the offspring are cousins, while genetically they are half-siblings (D'Onofrio et al., 2003; D'Onofrio et al., 2005; Gottesman & Bertelsen, 1989; Heath, Kendler, Eaves, & Markell, 1985; Nance & Corey, 1976; Rutter et al., 2001; Silberg & Eaves, 2004).

Cousin-comparisons enable researchers to examine risk factors that are almost always shared by siblings, like parental divorce, or by twins (e.g., SDP). The design also has different assumptions and limitations from sibling-comparisons, which makes them important because researchers need commensurate findings from multiple designs to draw strong causal inferences. Plus, cousin-comparisons can help researchers test assumptions in sibling-comparisons. For example, the comparison of differentially exposed first-born offspring of siblings/twins can be used to study a environmental risk free of the effects of birth order, which could confound sibling-comparison studies (Donovan & Susser, 2011; Lahey & D'Onofrio, 2010; Susser et al., 2010). Yet, cousin comparisons also have a number of limitations. Most importantly, the design cannot account for all genetic and environmental confounds, especially the

genetic and environmental influences passed down from the spouses of the twin parents (Eaves, Silberg, & Maes, 2005). When studying individual-level risk factors, such as SDP, the design, in fact, rules out fewer confounds than sibling-comparisons (e.g., cousin-comparisons do not account for as much genetic confounding as the comparison of full-sibling).

When offspring of multiple types of adult siblings are included in a study (e.g., offspring of full- *and* half-siblings and/or identical *and* fraternal twins) researchers can use quantitative-genetic models to specify the genetic and shared environmental sources of the familial confounding; that is, to estimate the extent to which environmental and genetic influences confound the statistical associations between risk factors and outcome (D'Onofrio et al., 2003; D'Onofrio et al., 2005; Gottesman & Bertelsen, 1989; Heath et al., 1985; Rutter et al., 2001; Silberg & Eaves, 2004). For instance, researchers can combine use offspring of siblings/twins to specify the familial background factors that confound sibling-comparisons (D'Onofrio et al., 2008; Harden et al., 2007; Kuja-Halkola et al., 2010). This is an important advantage because sibling-comparisons cannot identify the source of familial confounding when the comparison of differentially exposed siblings suggests no causal influence (Donovan & Susser, 2011; Lahey & D'Onofrio, 2010).

Adopted-at-birth Design

The children adopted-at-birth design can be used to explore whether the associations between prenatal factors and psychopathology are confounded by postnatal environmental factors. This is because the biological mothers provide genetic and prenatal environmental factors, but not the postnatal environment to the children. Thus, the design cannot remove effects of genetic confounding from prenatal environmental influences on children, such as SDP (Gaysina, Fergusson, Leve, et al., 2013).

By using a sample of children adopted-at-birth and their genetically unrelated rearing parents it is, on the other hand, possible to control for genetic confounding when examining the association between postnatal environmental factors (e.g., parenting) and child outcomes (Harold et al., 2013). This is because adoptive mothers provide postnatal environmental factors, but not genetic or prenatal environmental factors to children

In Vitro Fertilization Design

The *in vitro* fertilization (IVF) design is a special type of adoption study that relies new reproductive technologies to conduct "cross-fostering" studies. The design typically includes both offspring who are genetically related to mothers (e.g., when IVF is conducted on a woman's own egg) and not genetically related to their birth mothers (e.g., when there was an embryo donation) (Thapar et al., 2007). The design enables researcher to test whether a putative risk is associated with an offspring outcome in families where the offspring is not related to mother—the design rules out passive rGE. When IVF studies include offspring who are genetically related to their mothers *and* offspring who are not genetically related, researchers can explore the extent to which the statistical association between the risk and outcome is due to genetic confounding.(Thapar et al., 2007) Although the design has high internal validity,

there are concerns about the generalizability of findings from a sample of women who have had IVF (e.g., how representative are women who become pregnant through IVF and then smoke during pregnancy?).

Family-Based Quasi-Experimental Studies of SDP and ADHD

Clinical and epidemiological studies based on unrelated individuals show a consistent association (pooled analyses estimate the odds ratio as 2.36) and also dose–response relationships between maternal SDP and offspring ADHD (Langley, Rice, van den Bree, & Thapar, 2005b). In these studies, the associations between SDP and ADHD remained after controlling for measured covariates (e.g., maternal age at childbirth and parental psychopathology), which may therefore indicate a causal effect (Banerjee, Middleton, & Faraone, 2007; Langley, Rice, van den Bree, & Thapar, 2005a; Linnet et al., 2003; Motlagh et al., 2010; Rodriguez & Bohlin, 2005a). However, given that it is impossible to identify or accurately measure all the potential confounders, unmeasured confounding might also explain the observed associations.

Sibling-comparison studies of SDP and ADHD (D'Onofrio et al., 2008; Obel et al., 2011) have all reached the opposite conclusion—familial unmeasured familial factors account for all of the association. Although there is a robust correlation between maternal SDP and ADHD in studies using unrelated individuals, differentially exposed siblings do not differ in the rate of ADHD. There were concerns about the outcome measurement in some early sibling-comparison studies because they were based on maternal report or extreme cases(Rutter, 2007; Talati & Weissman, 2010), but all published sibling-comparison studies of ADHD to date, including studies predicting continuous trait measures (e.g., parent-ratings of offspring ADHD) and categorical measures (e.g., clinically diagnosed) of ADHD, have concluded that familial selection factors, not exposure to maternal SDP, account for the associations. These sibling-comparison results have been questioned by researchers who have explicitly hypothesized that women who vary in their smoking status across pregnancies are not comparable to all smoking women (Talati & Weissman, 2010). To address concerns about the generalizability of the findings from previous sibling-comparison and limitations inherent in the design (Talati & Weissman, 2010), a recent study included full-siblings, as well as full-cousins (Skoglund, Chen, D'Onofrio, Lichtenstein, & Larsson, 2013). This study found that the association between maternal SDP and ADHD gradually attenuated towards the null when adjusting for measured confounders, unmeasured confounders shared within the extended family (i.e., cousin comparison), and unmeasured confounders within the nuclear family (i.e., sibling comparison), suggesting that familial confounding accounts for the association between SDP and ADHD. Those results are also consistent with an offspring of twins study of SDP and offspring attention-deficit/hyperactivity problems (Knopik et al.). Thus, previous research suggest converging evidence across sibling and cousin comparison (offspring of full-siblings and offspring of twins), which provide strong support for the familiar confounding hypothesis.

A recent adoption study that suggests that postnatal genetic factors (e.g., the children's genetic factors influencing the adoptive parents behaviors) do not account for the association between SDP and offspring conduct problems (Gaysina et al., 2013).

The only available IVF study on SDP and offspring ADHD likewise suggests that genetic factors passed down from women account for the observed statistical associations (Thapar et al., 2009). This study included both offspring who were genetically related to (i.e., the mother's own egg was fertilized) and not genetically related to their birth mothers (i.e., the birth mother became pregnant through an embryo donation). SDP was significantly associated with offspring ADHD in the genetically related mother-offspring pairs, but not in the genetically unrelated mother-offspring pairs, which suggest inherited factors explain the associations. The importance of unmeasured familial factors is also supported by a recent study that compared the offspring ADHD risk between maternal and paternal SDP (Langley, Heron, Smith, & Thapar, 2012). Offspring ADHD symptoms were associated with exposure to both maternal and paternal SDP, which do not support a causal inference. Furthermore, when paternal smoking was examined in the absence of maternal smoking the associations remained and did not appear to be due to passive smoking exposure in utero. These findings together suggest that the association between maternal SDP and child ADHD may be due to unmeasured familial confounding. The converging evidence from studies using multiple quasi-experimental designs, therefore, provides strong evidence that unmeasured familial factors, including genetic confounds, account for the statistical associations between SDP and offspring ADHD.

Implications of Quasi-Experimental Research for Developmental Psychopathology

Family-based quasi-experimental designs have profound implications for developmental psychopathology research. First, quasi-experimental studies have important ramifications for research studying putative risk factors using epidemiological or convenience samples. It has become clear from recent quasi-experimental research, including studies that are based on comparing differentially exposed family members, that relies solely on including measured covariates in an analysis to control for confounds does not rule out all plausible alternative hypotheses. We encourage researchers to follow the scientific approach of identifying plausible alternative explanations for statistical associations ahead of time and using multiple designs to test competing causal hypotheses (Rutter et al., 2001; Shadish et al., 2002). This framework will often require researchers to stop the standard practice of conducting observational designs using one person from each family.

Second, quasi-experimental research has profound implications for basic science research on biological mechanisms related to psychopathology, an example of translational epidemiology (Hiatt, 2010; Talge, Neal, & Glover, 2007; Weissman, Brown, & Talati, 2011). As we have argued elsewhere (D'Onofrio & Lahey, 2010; D'Onofrio, Lahey, Turkheimer, & Lichtenstein, 2013), family-based quasi-experimental studies can greatly inform which mechanisms basic research should be exploring.

Third, the use of family-based quasi-experimental designs has important implications for research exploring biological moderating factors, such as gene by environment interaction (G×E) (Dick, 2011; Duncan & Keller, 2011). Studies that explore the interactions among specific genes and putative environmental risk factors are based on strong causal assumptions about the environmental influence (Conley & Rauscher,

2010; Dick, 2011; Moffitt, Caspi, & Rutter, 2005). As we illustrated above, many putative environmental risk factors may not have a specific influence on psychopathology; rather, the associations are due to confounding factors. As such, recent G × E studies of maternal SDP and psychopathology (Becker, El-Faddagh, Schmidt, Esser, & Laucht, 2008; L. S. Wakschlag et al., 2010), therefore, may have incorrectly interpreted the statistical interaction. Instead of a G × E interaction the statistical interaction could represent gene-by-gene interactions.

Finally, quasi-experimental designs are of great value in designing and testing preventions/interventions, a key aim of the field. Designing the best scientifically based interventions and prevention efforts relies on etiological studies that support strong causal inference regarding environmental exposures (Cicchetti, 1993; Coie et al., 1993; National Research Council & Institute of Medicine, 2009). Family-based quasi-experimental research can specify which environmental risk factors are associated with psychopathology independent of many environmental and genetic confounds. Interventions and preventative efforts, thus, aimed at decreasing these risk factors (or increasing these protection factors) could subsequently be studied using randomized controlled studies. It is also quite important to emphasize that quasi-experimental research can also identify risk factors that do not have a causal influence on psychopathology.

Summary

Researchers have identified environmental risk factors that are correlated with psychopathology. Based on these findings, researchers have developed complex developmental models and have tested biological moderating factors for these outcomes. We still know little about the true causes of psychopathology, however, because the extent to which the environmental risk factors are truly causal or due to familial confounding remains unclear (Academy of Medical Sciences Working Group, 2007; British Academy of Science Working Group, 2010). It is important to stress that this is true of both social science research, which has typically focused on environmental risk factors, and genetic research with susceptibility genes or other biomarkers (Turkheimer, 2012).

In this paper, we describe how family-based quasi-experimental methods can help inform our understanding of the true environmental causes of psychopathology. We argue that studies employing quasi-experimental methods play a unique position in bridging gaps between basic research and social science research because the approaches rigorously test causal hypotheses (Gaziano, 2010; Hiatt, 2010; Khoury, Gwinn, & Ioannidis, 2010; Weissman et al., 2011). We exemplified these principles using studies on maternal SDP and ADHD.

Acknowledgment

The writing of this chapter was supported by grants from the National Institute of Child Health and Human Development (HD061817 and HD061384), Swedish Research Council (Medicine), and Swedish Prison and Probation Services.

References

Academy of Medical Sciences Working Group. (2007). *Identifying the Environmental Causes of Disease: How should we decide what to believe and when to take action?* London: Academy of Medical Sciences.

Agrawal, A., Knopik, V. S., Pergadia, M. L., Waldron, M., Bucholz, K. K., Martin, N. G., et al. (2008). Correlates of cigarette smoking during pregnancy and its genetic and environmental overlap with nicotine dependence. *Nicotine and Tobacco Research*, *10*, 567–578

Bale, T. L., Baram, T. Z., Brown, A. S., Goldstein, J. M., Insel, T. R., McCarthy, M. M., et al. (2010). Early life programming and neurodevelopmental disorders. *Biological Psychiatry*, *68*(4), 314–319.

Banerjee, T. D., Middleton, F., & Faraone, S. V. (2007). Environmental risk factors for attention-deficit hyperactivity disorder. *Acta Paediatr*, *96*(9), 1269–1274.

Barker, D. J. P. (1998). *Mothers, Babies and Health in Later Life* (2nd ed.). Edinburgh: Churchill Livingstone.

Becker, K., El-Faddagh, M., Schmidt, M. H., Esser, G., & Laucht, M. (2008). Interaction of dopamine transporter genotype with prental smoke exposure on ADHD symptoms. *Journal of Pediatrics*, *152*, 263–269.

Bell, R. Q., & Harper, L. V. (1997). *Child Effects on Adults.* Hillsdale NJ: Lawrence Erlbaum.

Beydoun, H., & Saftlas, A. F. (2008). Physical and mental health outcomes of prenatal maternal stress in human and animal studies: A review of recent evidence. *Paediatric and Perinatal Epidemiology*, *22*, 438–466.

Biederman, J., Petty, C. R., Monuteaux, M. C., Fried, R., Byrne, D., Mirto, T., et al. (2010). Adult psychiatric outcomes of girls with attention deficit hyperactivity disorder: 11-year follow-up in a longitudinal case-control study. *The American Journal of Psychiatry*, *167*(4), 409–417.

British Academy of Science Working Group. (2010). *Social Science and Family Policy.* London: British Academy Policy Center.

Buss, C., Davis, E. P., Muftuler, L. T., Head, K., & Sandman, C. A. (2010). High pregnancy anxiety during mid-gestation is associated with decreased gray matter density in 6-9-year-old children. *Psychoneuroendocrinology*, *35*, 141–153.

Cicchetti, D. (1993). Developmental psychopathology: Reactions reflections projections. *Developmental Review*, *13*, 471–502.

Cnattingius, S. (2004). The epidemiology of smoking during pregnancy: Smoking prevalence, maternal characteristics, and pregnancy outcomes. *Nicotine and Tobacco Research*, *6*, S125–S140.

Coie, J. D., Watt, N. F., West, S. G., Hawkins, J. D., Asarnow, J. R., Markman, H. J., et al. (1993). The science of prevention: A conceptual framework and some directions for a national research program. *American Psychologist*, *48*, 1013–1022.

Conley, D., & Rauscher, E. (2010). *Genetic Interactions with Prenatal Social Environment: Effects on Academic and Behavioral Outcomes.* Cambridge, MA: National Bureau of Economic Research.

Cornelius, M. D., Ryan, C. M., Day, N. L., Godschmidt, L., & Willford, J. A. (2001). Prenatal tobacco effects on neuropsychological outcomes among preadolescents. *Journal of Developmental and Behavioral Pediatrics*, *22*, 217–225.

D'Onofrio, B. M., & Lahey, B. B. (2010). Biosocial influences on the family: A decade review. *Journal of Marriage and Family*, *72*, 762–782.

D'Onofrio, B. M., Lahey, B. B., Turkheimer, E., & Lichtenstein, P. (2013). Critical need for family-based, quasi-experimental designs in integrating genetic and social science research. *American Journal of Public Health*, *103* Suppl 1, S46–55.

D'Onofrio, B. M., Lahey, B. B., Turkheimer, E., & Lichtenstein, P. (2013). The critical need for family-based, quasi-experimental research in integrating genetic and social science research. *American Journal of Public Health*, *103*, S46–S55.

D'Onofrio, B. M., Turkheimer, E., Eaves, L. J., Corey, L. A., Berg, K., Solaas, M. H., et al. (2003). The role of the Children of Twins design in elucidating causal relations between parent characteristics and child outcomes. *Journal of Child Psychology & Psychiatry*, *44*, 1130–1144.

D'Onofrio, B. M., Turkheimer, E. N., Emery, R. E., Slutske, W., Heath, A., Madden, P. A. F., et al. (2005). A genetically informed study of marital instability and its association with offspring psychopathology. *Journal of Abnormal Psychology*, *114*, 570–586.

D'Onofrio, B. M., Van Hulle, C. A., Waldman, I. D., Rodgers, J. L., Harden, K. P., Rathouz, P. J., et al. (2008). Smoking during pregnancy and offspring externalizing problems: An exploration of genetic and environmental confounds. *Development and Psychopathology*, *20*, 139–164.

Dick, D. M. (2011). Gene-environment interaction in psychological traits and disorders. *Annual Review of Clinical Psychology*, *7*, 383–409.

Donovan, S. J., & Susser, E. (2011). Commentary: Advent of sibling designs. *International Journal of Epidemiology*, *40*, 345–349.

Duncan, G. J. (2012). Give us this day our daily breadth. *Child Development*, *83*(1), 6–15.

Duncan, L. E., & Keller, M. C. (2011). A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *American Journal of Psychiatry*, *168*, 1041–1049.

Eaves, L. J., Last, K., Martin, N. G., & Jinks, J. L. (1977). A progressive approach to non-additivity and genotype-environmental covariance in the analysis of human differences. *British Journal of Mathematical and Statistical Psychology*, *30*, 1–42.

Eaves, L. J., Last, K. A., Young, P. A., & Martin, N. G. (1978). Model-fitting approaches to the analysis of human behaviour. *Heredity*, *41*, 249–320.

Eaves, L. J., Silberg, J. L., & Maes, H. H. (2005). Revisiting the children of twins: Can they be used to resolve the environmental effects of dyadic parental treatment on child behavior? *Twin Research*, *8*, 283–290.

Ellingston, J. M., Rickert, M. E., Lichtenstein, P., Långström, N., & D'Onofrio, B. M. (2012). Disentangling the relationships between maternal smoking during pregnancy and co-occurring risk factors. *Psychological Medicine*, *42*, 1547–1557.

Ernst, M. (2001). Behavioral and neural consequences of prental exposure to nicotine. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*, 630–642.

Faraone, S. V., Biederman, J., & Mick, E. (2006). The age-dependent decline of attention deficit hyperactivity disorder: a meta-analysis of follow-up studies. *Psychological Medicine*, *36*(2), 159–165.

Faraone, S. V., Perlis, R. H., Doyle, A. E., Smoller, J. W., Goralnick, J. J., Holmgren, M. A., et al. (2005). Molecular genetics of attention-deficit/hyperactivity disorder. *Biological Psychiatry*, *57*(11), 1313–1323.

Fergusson, D. M. (1999). Prenatal smoking and antisocial behavior. *Archives of General Psychiatry*, *56*, 223–224.

Ferriero, D. M., & Dempsey, D. A. (1999). Impact of addictive and harmful substances on fetal brain development. *Current Opinion in Neurology*, *12*(2), 161–166.

Freese, J. (2008). Genetics and the social science explanation of individual outcomes. *American Journal of Sociology*, *114*, S1–S35.

Frisell, T., Oberg, S., Kuja-Halkola, R., & Sjolander, A. (2012). Sibling comparison designs: Bias from non-shared confounders and measurement error. *Epidemiology*, *23*, 713–720.

Froehlich, T. E., Lanphear, B. P., Auinger, P., Hornung, R., Epstein, J. N., Braun, J., et al. (2009). Association of tobacco and lead exposures with attention-deficit/hyperactivity disorder. *Pediatrics*, *124*(6), e1054–1063.

Galera, C., Cote, S. M., Bouvard, M. P., Pingault, J. B., Melchior, M., Michel, G., et al. (2011). Early risk factors for hyperactivity-impulsivity and inattention trajectories from age 17 months to 8 years. *Archives of General Psychiatry*, *68*(12), 1267–1275.

Gaysina, D., Fergusson, D. M., Leve, L. D., et al. (2013). Maternal smoking during pregnancy and offspring conduct problems: Evidence from 3 independent genetically sensitive research designs. *JAMA Psychiatry*, *70*(9), 956–963.

Gaziano, J. M. (2010). The evolution of population science. *JAMA: The Journal of the American Medical Association*, *304*(20), 2288–2289.

Gilman, S. E., & Loucks, E. B. (2012). Invited commentary: Does the Childhood environment influence the association between every *x* and every *y* in adulthood? *American Journal of Epidemiology*, *176*(8), 684–688.

Glantz, M. D., & Chambers, J. C. (2006). Prenatal drug exposure effects on subsequent vulnerability to drug abuse. *Development and Psychopathology*, *18*(3), 893–922.

Glantz, M. Y., & Campbell Champers, J. (2006). Prenatal drug exposure effects on subsequent vulnerability to drug abuse. *Development and Psychopathology*, *18*, 893–922.

Gluckman, P. D., & Hanson, M. A. (2007). Developmental plasticity and human disease: Research directions. *Journal of Internal Medicine*, *261*(5), 461–471.

Gluckman, P. D., Hanson, M. A., Cooper, C., & Thornburg, K. L. (2008). Effect of *in utero* and early-life conditions on adult health and disease. *New England Journal of Medicine*, *359*, 61–73.

Gottesman, I. I., & Bertelsen, A. (1989). Confirming unexpressed genotypes for schizophrenia. *Archives of General Psychiatry*, *46*, 867–872.

Halperin, J. M., & Schulz, K. P. (2006). Revisiting the role of the prefrontal cortex in the pathophysiology of attention-deficit/hyperactivity disorder. *Psychological bulletin*, *132*(4), 560–581.

Harden, K. P., Lynch, S. K., Turkheimer, E., Emery, R. E., D'Onofrio, B. M., Slutske, W. S., et al. (2007). A behavior genetic investigation of adolescent motherhood and offspring mental health problems. *Journal of Abnormal Psychology*, *116*, 667–683.

Harold, G. T., Leve, L. D., Barrett, D., Elam, K., Neiderhiser, J. M., Natsuaki, M. N., et al. (2013). Biological and rearing mother influences on child ADHD symptoms: revisiting the developmental interface between nature and nurture. *Journal of Child Psychology and Psychiatry*, *54*(10), 1038–1046.

Heath, A. C., Kendler, K. S., Eaves, L. J., & Markell, D. (1985). The resolution of cultural and biological inheritance: Informativeness of different relationships. *Behavior Genetics*, *15*, 439–465.

Heath, A. C., Kessler, R. C., Neale, M. C., Hewitt, J. K., Eaves, L. J., & Kendler, K. S. (1993). Testing hypotheses about direction of causation using cross-sectional data. *Behavior Genetics*, *23*, 29–50.

Hiatt, R. A. (2010). Invited commentary: The epicenter of translational science. *American Journal of Epidemiology*, *172*(5), 525–527.

Huizink, A. C., & Mulder, E. J. H. (2006). Maternal smoking, drinking or cannabis use during pregnancy and neurobehavioral and cognitive functioning in human offspring. *Neuroscience and Biobehavioral Reviews 30* 24–41.

Huizink, A. C., Robles de Medina, P. G., Mulder, J. H., Visser, G. H. A., & Buitelaar, J. K. (2003). Stress during pregnancy is associated with developmental outcome in infancy. *Journal of Child Psychology & Psychiatry*, *44*(6), 810–818.

Iliadou, A. N., Koupil, I., Villamor, E., Altman, D., Hultman, C., Långström, N., et al. (2010). Familial factors confound the association between maternal smoking during pregnancy and young adult offspring overweight. *International Journal of Epidemiology*, *39*, 1193–1202.

Jaffee, S. R., & Price, T. S. (2012). The implications of genotype–environment correlation for establishing causal processes in psychopathology. *Development and Psychopathology*, *24*(Special Issue 04), 1253–1264.

Kendler, K. S. (2005). Psychiatric genetics: A methodological critique. *American Journal of Psychiatry*, *162*, 3–11.

Kendler, K. S., & Baker, J. H. (2007). Genetic influences on measures of the environment: A systematic review. *Psychological Medicine*, *37*, 615–626.

Khoury, M. J., Gwinn, M., & Ioannidis, J. P. A. (2010). The emergence of translational epidemiology: From scientific discovery to population health impact. *American Journal of Epidemiology*, *172*(5), 517–524.

Knopik, V. S., Sparrow, E. P., Madden, P. A. F., Bucholz, K. K., Hudziak, J. J., Reich, W., et al. (2005). Contributions of parental alcoholism, prenatal substance exposure, and genetic transmission to child ADHD risk: A female twin study. *Psychological Medicine*, *35*, 625–635.

Koob, G. F. (1999). Stress, corticotropin releasing factor, and drug addiction. *Annals of the New York Academy of Sciences*, *897*(1), 27–45.

Kraemer, H. C., Kazdin, A. E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupfer, D. J. (1997). Coming to terms with the terms of risk. *Archives of General Psychiatry*, *54*, 337–343.

Kramer, M. S. (2000). Inivited commentary: Association between restricted fetal growth and adult chronic disease: Is it causal? Is it important? *American Journal of Epidemiology*, *152*(7), 605–608.

Kuja-Halkola, R., D'Onofrio, B. M., Illiadou, A., Pawitan, Y., Langstrom, N., & Lichtenstein, P. (2010). Prenatal smoking exposure and stress coping in late adolescence: No causal link. *International Journal of Epidemiology*, *39*, 1531–1540.

Lahey, B. B., & D'Onofrio, B. M. (2010). All in the family: Comparing siblings to test causal hypotheses regarding environmental influences on behavior. *Current Directions in Psychological Science.*, *19*, 319–323.

Lahey, B. B., D'Onofrio, B. M., & Waldman, I. D. (2009). Using epidemiologic methods to test hypotheses regarding causal influences on child and adolescent mental disorders. *Journal of Child Psychology and Psychiatry*, *50*, 53–62.

Lambe, M., Hultman, C., Torrang, A., MacCabe, J., & Cnattingius, S. (2006). Maternal smoking during pregnancy and school performance at age 15. *Epidemiology*, *17*, 524–530.

Langley, K., Heron, J., Smith, G. D., & Thapar, A. (2012). Maternal and paternal smoking during pregnancy and risk of ADHD symptoms in offspring: Testing for intrauterine effects. *American Journal of Epidemiology*, *176*(3), 261–268.

Langley, K., Rice, F., van den Bree, M. B., & Thapar, A. (2005a). Maternal smoking during pregnancy as an environmental risk factor for attention deficit hyperactivity disorder behaviour. *A review. Minerva Pediatr*, *57*(6), 359–371.

Langley, K., Rice, F., van den Bree, M. B., & Thapar, A. (2005b). Maternal smoking during pregnancy as an environmental risk factor for attention deficit hyperactivity disorder behaviour. *A review. Minerva pediatrica*, *57*(6), 359–371.

Larsson, H., Anckarsater, H., Rastam, M., Chang, Z., & Lichtenstein, P. (2012). Childhood attention-deficit hyperactivity disorder as an extreme of a continuous trait: a quantitative genetic study of 8,500 twin pairs. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *53*(1), 73–80.

Lawlor, D. A., & Mishra, G. D. (2009). *Family Matters: Designing, Analysing, and Understanding Family-Based Studies in Life Course Epidemiology*. Oxford: Oxford University Press.

Linnet, K. M., Dalsgaard, S., Obel, C., Wisborg, K., Henriksen, T. B., Rodriguez, A., et al. (2003). Maternal lifestyle factors in pregnancy risk of attention deficit hyperactivity disorder and associated behaviors: review of the current evidence. *American Journal of Psychiatry*, *160*(6), 1028–1040.

Lotfipour, S., Ferguson, E., Leonard, G., Perron, M., Pike, B., Richer, L., et al. (2009). Orbitofrontal cortex and drug use during adolescence: Role of prenatal exposure to maternal smoking and BDNF genotype. *Archives of General Psychiatry*, *66*(11), 1244–1252.

Lundberg, F., Cnattingius, S., D'Onofrio, B., Altman, D., Lambe, M., Hultman, C., et al. (2010). Maternal smoking during pregnancy and intellectual performance in young adult Swedish male offspring. *Pediatric and Perinatal Epidemiology*, *24*, 79–87.

Maughan, B., Taylor, A., Caspi, A., & Moffitt, T. E. (2004). Prenatal smoking and early childhood conduct problems. *Archives of General Psychiatry*, *61*, 836–843.

McGue, M., Osler, M., & Christensen, K. (2010). Causal inference and observational research: The utility of twins. *Perspectives on Psychological Science*, *5*, 546–556.

Meyer, K. A., Williams, P., Hernandez-Diaz, S., & Cnattingius, S. (2004). Smoking and the risk of oral clefts: Exploring the impact of study designs. *Epidemiology*, *15*, 671–678.

Mick, E., Biederman, J., Faraone, S. V., Sayer, J., & Kleinman, S. (2002). Case control study of attention deficit hyperactivity disorder and maternal smoking, alcohol use, and drug use during pregnancy. *Journal of the American Academy of Child and Adolescent Psychiatry*, *41*, 378–385.

Milberger, S., Biederman, J., Faraone, S. V., & Jones, J. (1998). Further evidence of an association between maternal smoking during pregnancy and attention deficit hyperactivity disorder: findings from a high-risk sample of siblings. *Journal of Clinical Child Psychology*, *27*(3), 352–358.

Moffitt, T. E., Caspi, A., & Rutter, M. (2005). Strategy for investigating interactions between measured genes and measured environments. *Archives of General Psychiatry*, *62*, 473–481.

Motlagh, M. G., Katsovich, L., Thompson, N., Lin, H., Kim, Y. S., Scahill, L., et al. (2010). Severe psychosocial stress and heavy cigarette smoking during pregnancy: an examination of the pre- and perinatal risk factors associated with ADHD and Tourette syndrome. *European Child and Adolescent Psychiatry*, *19*(10), 755–764.

Nance, W. E., & Corey, L. A. (1976). Genetic models for the analysis of data from the families of identical twins. *Genetics*, *83*, 811–826.

National Research Council & Institute of Medicine. (2009). *Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and Possibilities*. Washington, DC: National Academies Press.

Neale, M. C., & Cardon, L. R. D. (1992). *Methodology for Genetic Studies of Twins and Families*. Dordrecht, The Netherlands: Kluwer Academic Press.

O'Connor, T. G. (2003). Natural experiments to study the effects of early experience: Progress and limitations. *Development and Psychopathology*, *15*, 837–852.

O'Connor, T. G., Heron, J., Golding, J., Beveridge, M., & Glover, V. (2002). Maternal antenatal anxiety and children's behavioural/emotional problems at 4 years. *British Journal of Psychiatry*, *180*, 502–508.

O'Connor, T. G., Heron, J., Golding, J., Glover, V., & ALSPAC study team. (2003). Maternal antenatal anxiety and behavioural/emotional problems in children: a test of a programming hypothesis. *Journal of Child Psychology & Psychiatry*, *44*(7), 1025–1036.

O'Connor, T. G. (2003). Natural experiments to study the effects of early experience: Progress and limitations. *Development and Psychopathology*, *15*, 837–852.

Obel, C., Linnet, K. M., Henriksen, T. B., Rodriguez, A., Jarvelin, M. R., Kotimaa, A., et al. (2009). Smoking during pregnancy and hyperactivity-inattention in the offspring–comparing results from three Nordic cohorts. *International Journal of Epidemiology*, *38*(3), 698–705.

Obel, C., Olsen, J., Henriksen, T. B., Rodriguez, A., Järvelin, M.-R., Moilanen, I., et al. (2011). Is maternal smoking during pregnancy a risk factor for hyperkinetic disorder?—findings from a sibling design. *International Journal of Epidemiology*, *40*, 338–345.

Plomin, R., & Bergeman, C. S. (1991). The nature of nurture: Genetic influences on "environmental" measures. *Behavioral and Brain Sciences*, *10*, 1–15.

Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype–environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin*, *84*, 309–322.