Summer 8-10-2021

Factors Influencing the Accuracy of Estimates in Hierarchical Age-Period-Cohort Models: A Monte Carlo Simulation Study

Brandon Attell

ACCEPTANCE

This dissertation, FACTORS INFLUENCING THE ACCURACY OF ESTIMATES IN HIERARCHICAL AGE-PERIOD-COHORT MODELS: A MONTE CARLO SIMULATION STUDY, by BRANDON ATTELL, was prepared under the direction of the candidate's Dissertation Advisory Committee. It is accepted by the committee members in partial fulfillment of the requirements for the degree, Doctor of Philosophy, in the College of Education & Human Development, Georgia State University.

The Dissertation Advisory Committee and the student's Department Chairperson, as representatives of the faculty, certify that this dissertation has met all standards of excellence and scholarship as determined by the faculty.

Audrey J. Leroux, Ph.D.
Committee Chair

Hongli Li, Ph.D.
Committee Member

C. Kevin Fortner, Ph.D.
Committee Member

Eric Wright, Ph.D.
Committee Member

Date

Jennifer Esposito, Ph.D.
Chairperson, Department of Educational Policy Studies

Paul A. Alberto, Ph.D.
Dean, College of Education & Human Development

AUTHOR'S STATEMENT

NOTICE TO BORROWERS

All dissertations deposited in the Georgia State University library must be used in accordance with the stipulations prescribed by the author in the preceding statement. The author of this dissertation is:

Brandon Kyle Attell
Educational Policy Studies
College of Education & Human Development
Georgia State University

The director of this dissertation is:

Audrey J. Leroux, Ph.D.
Department of Educational Policy Studies
College of Education and Human Development
Georgia State University
Atlanta, GA 30303

CURRICULUM VITAE

Brandon Attell

ADDRESS: 30 Pryor Street NW
Atlanta, GA 30303

EDUCATION:

Ph.D.	2021	Georgia State University Educational Policy Studies Research, Measurement, and Statistics
M.A.	2015	Georgia State University Sociology
B.S.	2012	Kennesaw State University Sociology

PROFESSIONAL EXPERIENCE:

2018-present	Scientific Data Analyst Cadence Group, TekSystems, and Chenega
2016-2018	Research Associate Georgia Health Policy Center
2012-2016	2CI Research Fellow Georgia State University Department of Sociology

PRESENTATIONS AND PUBLICATIONS:

Hatfield, K. M., Reddy, S. C., Forsberg, K., Korhonen, L., Garner, K., Gulley, T., ... & Sievers, M. (2020). Facility-wide testing for SARS-CoV-2 in nursing homes-Seven US jurisdictions, March-June 2020. *MMWR. Morbidity and Mortality Weekly Report*, *69*(32), 1095-1099.

Attell, B.K. (2020). Changing attitudes toward euthanasia and suicide for terminally ill persons, 1977-2016: An age-period-cohort analysis. *Journal of Death and Dying*, *80*, 355-379.

Attell, B.K., Cappelli, C., Manteuffel, B, & Li, H. (2020). Measuring functional impairment in children and adolescents: Psychometric properties of the Columbia Impairment Scale (CIS). *Evaluation & the Health Professions, 43,* 3-15.

Minyard, K., Manteuffel, B., Smith, C.M., Attell, B.K., Landers, G., Schlanger, M., & Dore, E. (2019). *Treatment services for people with co-occurring substance use and mental health problems: A rapid realist synthesis*. Dublin, Ireland: Ireland Health Research Board. 96 Pages.

Attell, B.K., Kummerow, K., & Treiber, L.A. (2017). Workplace bullying, perceived job stressors, and psychological distress: Gender and race differences in the stress process. *Social Science Research, 65,* 210-221.

Wright, E.R., Attell, B.K., & Ruel, E.E. (2017). Social support networks and the mental health of runaway and homeless youth. *Social Sciences*, *6*, 1-16.

Attell, B.K. (2016). Disparities in health insurance coverage among children and young adults in Georgia and the U.S. *Journal of the Georgia Public Health Association, 6,*155-157.

Treiber, L.A., Markle, G.L., & Attell, B.K. (2016). Cascading losses: A case study of suffering from AIDS and head and neck cancer. *Illness, Crisis & Loss, 24,* 217-235.

Hatch, A.R., Xavier-Brier, M., Attell, B.K., & Grucza-Viscara, E. (2016). Soldier, elder, prisoner, ward: Psychotropics in the era of transinstitutionalization. *Advances in Medical Sociology, 17,* 291-317.

Markle, G.L., Attell, B.K., & Treiber, L.A. (2015). Dual, yet dueling illnesses: Multiple chronic illness experience at midlife. *Qualitative Health Research,* 25, 1271-1282.

Attell, B.K. (2014). Book review: Dancing at the river's edge: A patient and her doctor negotiate life with chronic illness. *The Journal of Public and Professional Sociology, 6,* 1-5.

Attell, B.K. (2013). Social contact theory: A framework for understanding AIDS-related stigma. *The Journal of Public and Professional Sociology*, *5,* 1-23.

Attell, B.K. (2011). Book review: Using women: Gender, drug policy, and social justice. *The Journal of Public and Professional Sociology, 3*, 1-7.

PROFESSIONAL SOCIETIES AND ORGANIZATIONS

2011	American Sociological Association
2016	American Statistical Association

FACTORS INFLUENCING THE ACCURACY OF ESTIMATES IN HIERARCHICAL AGE-PERIOD-COHORT MODELS: A MONTE CARLO SIMULATION STUDY

by

BRANDON ATTELL

Under the Direction of Audrey J. Leroux, Ph.D.

ABSTRACT

Social scientists have long been interested in the role that age, period, and cohort effects have in influencing longitudinal trends in a variety of research areas. However, because the three effects are linear derivatives of one another traditional statistical models cannot simultaneously isolate their unique effects due to their perfect confounding, an issue known as the identification problem. One recent solution to the identification problem is the estimation of age, period, and cohort effects through the use of cross-classified random effects modeling applied to repeated cross-sectional data. This approach takes advantage of the multilevel modeling framework and proposes that age can be treated as an individual-level (level-one) variable and period and cohort effects can be treated as categorical variables that define cluster membership at level-two, allowing all three effects to be simultaneously estimated. Using a Monte Carlo simulation study, this dissertation investigated two broad areas of methodological issues related to the performance of the model. First, four factors were manipulated that may influence the accuracy of estimates in the model: the number of survey years available for analysis; the cohort grouping employed in the model; the variability of the period effect; and the variability of the cohort effect. Second, four model fit indices were evaluated to determine their performance in detecting the cohort grouping underlying the structure of the dataset used in the analysis. The results of the simulation study indicated that the cohort grouping used in the model heavily influenced the accuracy of many of the model parameters, while the number of survey years available for analysis most directly influenced the accuracy of the period-level predictor. To a lesser extent, variability in the period and cohort effects impacted the accuracy of a few of the model parameters, but tended to occur in scenarios where bias was exhibited due to the cohort grouping. Importantly, all four of the model fit indices performed well in detecting the cohort grouping underlying the dataset. Implications of these findings are discussed for applied researchers, funders and administrators of repeated cross-sectional surveys, and life course theorists. Areas for future methodological research are also provided.

INDEX WORDS: age-period-cohort, longitudinal research, life course theory, hierarchical linear models, Monte Carlo simulation

FACTORS INFLUENCING THE ACCURACY OF ESTIMATES IN HIERARCHICAL AGE-PERIOD-COHORT MODELS: A MONTE CARLO SIMULATION STUDY

by

BRANDON ATTELL

A Dissertation

Presented in Partial Fulfillment of Requirements for the

Degree of

Doctor of Philosophy

in

Research, Measurement, and Statistics

in

Educational Policy Studies

in

the College of Education & Human Development

Georgia State University

Atlanta, GA
2021

ACKNOWLEDGMENTS

I would first like to thank my dissertation chair, Dr. Audrey Leroux, for her guidance and input throughout this project. Our weekly meetings kept me on track and gave me something to look forward to throughout the process of completing the dissertation. Her expertise in cross-classified random effects modeling and Monte Carlo simulation studies greatly strengthened my study. I would also like to thank my committee members, Dr. Hongli Li, Dr. Kevin Fortner, and Dr. Eric Wright for their thoughtful comments and insights that helped to improve my work. My gratitude is also extended to the many family members, friends, coworkers, and colleagues who checked in on me from time to time and offered their encouraging words to see this through to the end. Finally, thank you to my partner Jacob, who has unrelentingly supported me more than anyone else … I finally finished!

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AIC	Akaike's Information Criterion
AICC	Finite Sample Corrected Akaike's Information Criterion
BIC	Bayesian Information Criterion
BICC	Cluster Size Corrected Bayesian Information Criterion
APC	Age-Period-Cohort
APC-C	Age-Period-Cohort Characteristics
C-APC	Classical Age-Period-Cohort
CCREM	Cross-Classified Random Effects Model
CG-APC	Constrained Generalized Age-Period-Cohort
GSS	General Social Survey
HAPC-CCFEM	Hierarchical Age-Period-Cohort Cross-Classified Fixed Effects Model
HAPC-CCREM	Hierarchical Age-Period-Cohort Cross-Classified Random Effects Model
REML-EB	Restricted Maximum Likelihood Estimation - Empirical Bayes

1 INTRODUCTION

Social scientists have long been interested in the study of social change. Indeed, cross-sectional research may be limited if the research design relies entirely on data collected at one point in time, resulting in conclusions that must be interpreted as correlational instead of causational (Kraemer et al., 2000). In these scenarios longitudinal research offers an advantage to cross-sectional studies by considering the influence of data collected at multiple points in time, laying the groundwork for the measurement of change, particularly the magnitude and direction among key variables and outcomes of interest (Podsakoff et al., 2011; Ruspini, 1999). As it relates to longitudinal research, social scientists importantly recognize and place emphasis on the fact that social change occurs not only among individuals but is also influenced by the broader social contexts and circumstances in which they are situated (Bernardi et al., 2019, 2020; Landes & Settersten, 2019). This leads to an understanding of social change as occurring at both the individual and societal level. One theoretical framework aligned to this particular conceptualization of social change and longitudinal research is life course sociological theory.

Life Course Theory

Proponents of life course theory contend that any longitudinal change in social phenomenon can be attributed to three competing and simultaneous forces: age effects, period effects, and cohort effects. Age effects refer to the situation in which longitudinal change in an outcome of interest can be partly or wholly attributed to the biological aging process that occurs as an individual gets older (Keyes & Li, 2012). Therefore, age effects are measured at the individual level, and they represent person-level changes over time as people move throughout the life course. For example, as individuals age they are at greater risk of death as their physical health deteriorates in later life. Similarly, risk for chronic conditions such as heart attack, stroke, or cancer are typically age-graded in which individuals are more likely to experience these events as they age biologically.

These age effects are not only relegated to developmental or biological processes. As individuals move throughout the life course, phenomena such as attitudes, beliefs, or ideas might change for a variety of reasons. For example, life course sociologists have pointed out the importance of life transitions that occur as we age. Life transitions refer to the important socially constructed roles that we take on throughout life that impact our attitudes and viewpoints, and include transitions such as graduating from college, getting married, and experiencing the death of parents (George, 1993). Life transitions are inherently tied to the "age structuring" of society (Settersten & Mayer, 1997). That is, societies have social and cultural expectations and norms that individuals typically conform to regarding important life accomplishments, roles, and positions that should be achieved at established time points during the life course. For example, one can argue that it is more socially acceptable to rely on parents for financial support during childhood, but less so by age 40 or 50. Finally, as individuals age they experience formative life events that impact their outlook on the world. Accordingly, differences in attitudes towards an outcome of interest may vary by age given these variations in life experiences (Elder, 1994, 1998).

As pointed out by Glenn, however, "age differences in a variable may reflect the effects of being born at different times and having different formative experiences rather than or in addition to the effects of growing older" (2002, p. 465). Therefore, while recognizing the important role that age plays in influencing longitudinal change over time, life course theorists also prioritize the need to situate individuals in the historical contexts to which they are associated, while also taking into account changing social contexts over time (Heinz & Kruger, 2001). Taking into account broader contextual forces is especially important because while individuals might exhibit their own agency and make their own choices, the opportunities and actions available to them may be constrained by societal forces and historical timelines (Elder, 1998). The first of these broader contextual forces is that of the cohort effect. A cohort can be defined as any group of individuals that collectively experience some event at the same time (Yang, 2011). For example, in public health research, it is common to define cohorts by delineating a group of individuals who collectively experience a particular disease or illness, such as the cohort of babies born with birth defects as a result of the 2015 Zika virus outbreak in Brazil (Satterfield-Nash et al., 2017). Once a cohort is conceptualized, it can be compared to other cohorts for differences in outcomes over time.

Perhaps the most commonly utilized cohort in the application of life course theory is the birth cohort, or all individuals who are born during a specified time period (Glenn, 2005). Life course sociologists have described several reasons why birth cohorts are particularly influential sources of variance in change over time, especially as they relate to attitudinal changes. One explanation is that what people think of current events and social phenomenon is equally explainable by not only what is currently taking place in the world, but what was also happening when individuals were growing up (Alwin & McCammon, 2003). For example, birth cohort differences regarding attitudes toward technology may exist between cohorts who spent much of their childhood without technology compared to cohorts that grew up in the proliferation of the Internet and personal devices like cellphones and tablets. From this point of view, the birth cohort is a useful analytical tool that indexes commonly shared experiences tied to a specific historical era that an individual carries with them throughout life (Alwin & McCammon, 2003). From a macro perspective, birth cohorts influence longitudinal trends as new cohorts come into existence, bringing new experiences and attitudes, while older cohorts die out of the population, taking their experiences and attitudes with them, a process life course sociologists have referred to as demographic metabolism and cohort succession (Firebaugh, 1997; Glenn, 2005; Lutz, 2012; Ryder, 1965).

In addition to cohort effects, life course sociologists contend that a second historical context must also be taken into consideration: period effects. The life course theoretical literature has conceptualized period effects in two ways. Perhaps most simply, some contend that period effects merely represent the date at which an outcome or observation is measured (Fosse & Winship, 2019). However, it is more commonly recognized that period effects can also be conceptualized as variations over time, typically calendar years, that influence all age groups simultaneously (Keyes & Li, 2012). In this view, period effects are important to consider because they subsume important historical events, environmental changes, and changing social and cultural contexts (Yang, 2011). As an example, a researcher examining yearly trends in attitudes toward gun violence might notice substantial peaks in anti-gun attitudes in years where mass school shootings occurred. Life course sociologists contend that period effects are therefore indicators that capture important events like the Columbine shooting or Sandy Hook shooting that simultaneously impact or expose all age groups to some phenomenon that may impact the outcome of interest, often times quantifying sudden institutional changes or shifts over years (Mayer, 2009). Stated another way, we can view period effects as explaining potential peaks and valleys exhibited in certain years of data for an outcome, and they can also represent changes due to the broad passage of time or the linear passage of calendar years that are thought to capture the many different "shifts in social, cultural, economic, or physical environments" (Yang & Land, 2013, p. 2) that may influence an outcome. The notion that period effects capture broad shifts over years is particularly important given that life course scholars are generally interested in examining trends over long periods of time, typically over several decades. For example, one such period effect is exhibited in changing attitudes towards same-sex marriage in the United States. In her analysis using data from the General Social Survey between 1988 and 2010, Baunach (2012) found strong period effects; namely that in 1988 only 3.3% of Americans strongly agreed to same-sex marriage, which increased to 21.6% in 2010.

In light of life course theory, a successful longitudinal analysis must take into account all three temporal aspects that comprise age, period, and cohort effects. Otherwise, conclusions regarding one of the effects, if examined in isolation of the others, may result in erroneous findings as to the true processes responsible for social change. An age-period-cohort (APC) model, therefore, is generally any model that attempts to estimate the unique contributions of age effects, period effects, and cohort effects at the same time (Costanza et al., 2017; Keyes & Li, 2012). These models are particularly useful in their alignment with life course theory in that they recognize and prioritize the "multilevel phenomenon, ranging from structured pathways through social institutions and organizations to the social trajectories of individuals and their developmental pathways" (Elder, 1994, p. 5) that each age, period, and cohort process might have in contributing to longitudinal change over time in an outcome of interest. That is, APC models attempt to not only examine individual-level change over time (age effects), but also equally consider the important roles that period effects and cohort effects have by linking individuals to these influential social contexts (Elder, 1998; Suzuki, 2012).

If an APC model appropriately accounts for all three temporal parameters, researchers can better understand which of the processes are more important in influencing a particular outcome. Such a better understanding has important implications for policy and program interventions. For example, Riether et al. (2009) conducted an APC analysis of trends in the prevalence of obesity in the United States from 1976 to 2002 using a sample of 1.7 million adults from the National Health Interview Surveys. The results of their APC analysis indicated relatively small age and cohort effects, but large period effects that were primarily responsible for rising rates of obesity. They concluded that the "secular trend" hypothesis regarding increases in the prevalence of obesity is supported by APC analysis, meaning that longitudinal changes in obesity are likely due to shifting yearly trends at the societal level in issues like more time spent working and less time for leisure and exercise activity. Therefore, intervening from a policy or programmatic perspective requires solutions that take into account not only individual-level factors, but societal-level factors as well. As such, APC models may help shed light on the most appropriate level to intervene at when addressing important and complex social problems.

The Identification Problem

Publicly available, longitudinal datasets like that used by Riether et al. (2009) capturing the variables necessary for APC models are ubiquitous. Therefore, the inherit challenge in APC modeling does not lie in data availability but is instead an estimation issue. Although statistical models aligned with life course theory have advanced over time through the implementation of growth curve models and event history analysis (Elder et al., 2003; Mayer, 2009), specific models attempting to uniquely separate longitudinal change in any outcome that is attributable to each distinct age, period, and cohort process continue to suffer from what has long been recognized as the identification problem (Blalock, 1967; Fienberg & Mason, 1979; K. O. Mason et al., 1973). Consider the following. If a survey is administered to an individual in the year 2018 and demographic information regarding the respondent's age is collected, say 20 years old, their birth cohort membership can be easily calculated by subtracting the respondent's age from the survey year. In this case, the respondent that is 20 years old in 2018 was born in 1998, giving them membership in the 1998 birth cohort. Alternatively, if in the same survey birth cohort information was collected for a respondent, say 1975, their age can be derived by subtracting the birth cohort from the survey year, in this case resulting in a respondent that is 43 years old. Finally, if birth cohort information and age in years at the time of the survey is available, the year in which a respondent was surveyed can be determined by adding the respondent's birth cohort and age together. Therefore, age, period, and cohort effects are linear derivatives of one another and can be represented as:

$$\begin{gathered} \text{Age} = \text{period} - \text{cohort}, \\ \text{Period} = \text{age} + \text{cohort}, \\ \text{Cohort} = \text{period} - \text{age}. \end{gathered} \tag{1}$$

The perfectly linear combinations of age, period, and cohort result in a statistical model that is fully specified if all three variables are entered into the same model at once. Therefore, the linear relationships between age, period, and cohort are problematic because traditional statistical models such as ordinary least squares regression or logistic regression cannot simultaneously isolate their unique effects due to their perfect confounding. This situation has been termed the "identification problem" and has plagued researchers for the past thirty to forty years. From a life course theoretical perspective, the estimation of all three effects simultaneously is imperative to more fully understand the exact mechanisms responsible for longitudinal changes over time. However, the identification problem makes the concurrent estimation of these effects mathematically impossible. Therefore, despite what some have termed a "futile quest" (Bell & Jones, 2014a; Glenn, 1976), sociologists, demographers, statisticians, and other researchers have devoted significant attention to finding solutions and workarounds to the APC identification problem.

Statement of Purpose

One recent solution to the identification problem that has been particularly appealing to life course scholars is the use of sophisticated multilevel modeling techniques to estimate age, period, and cohort effects simultaneously. As will be discussed in more detail in Chapter 2, multilevel models are used to analyze clustered data collected at different levels of measurement, for example students (level-one) nested within classrooms (level two). As it relates to solving the identification problem, Yang and Land (2006) proposed that age could be treated as an individual level variable (level one) and period and cohort effects could be treated as categorical variables that define cluster membership at level two. As such, they contend the identification problem could be solved by using a multilevel model to simultaneously estimate all three APC effects.

The use of such a model is proliferated in the applied literature, especially among life course scholars studying historical changes in physical and mental health, and social scientists examining changing attitudes towards a variety of social, political, and cultural issues. Yet, despite the popularity of the model in the applied literature, less attention has been given to the methodological development of the model beyond its initial articulation fifteen years ago. Using a Monte Carlo simulation study, the purpose of this dissertation is to fill this gap in the literature by examining several important methodological components of the model faced by researchers seeking to use it in practice. Broadly, there are two areas of investigation. The first area is concerned with the statistical accuracy of the estimates derived from the model. Previous simulation studies examined the influence that the number of survey years had on the accuracy of the model estimates and the ability of the model to accurately recover period and cohort trends. This dissertation expands on previous simulation studies by re-examining the impact of the number of repeated cross-sections on accuracy of estimates in the model in addition to three other factors: the cohort groupings employed in the model, the variability of the period effect, and the variability of the cohort effect.

The second area of investigation is concerned with how applied researchers should specify cohort groups in the model. In reality, birth cohorts are a continuous measure given that they are derived from an individual's year of birth. However, to define level-two cluster membership in a multilevel model they must be transformed and treated as categorical variables that define an individual's membership into a level-two cohort grouping. Common cohort groupings used in previous research include the three-year cohort, five-year cohort, and ten-year cohort. While any cohort grouping could be arbitrarily chosen, many applied researchers contend that a balance should exist between specificity and breadth that exists in the heterogeneity of individuals captured in a particular cohort grouping, with the five-year cohort serving as an ideal balance that is used in many models in the applied literature. Therefore, the choice of which cohort grouping to use is an important decision faced by applied researchers. While theoretically grounded, no methodological attention has been devoted to examining how applied researchers can define such cohort groups in the model. One potential is the use of model fit indices to determine which cohort grouping best fits the underlying structure of the data in the analysis. Therefore, this dissertation aims to examine if model fit indices could work in this way and to identify which commonly used fit indices can aid applied researchers in determining the best cohort size to use in the model.

2 REVIEW OF THE LITERATURE

This chapter provides a review of the literature for estimating age-period-cohort (APC) models and describes the purpose of the current study. The review begins by examining several historical methods of estimating APC models that attempt to resolve or work around the identification issue described in Chapter 1. As pointed out by Fosse and Winship (2019), hundreds or even thousands of articles exist debating various solutions to the identification issue. Therefore, the review provides an overview of several of the most commonly utilized solutions to the APC identification problem. These solutions were specifically selected for review as they were consistently identified by other researchers in several different lines of previous research tracing the historical development of estimating APC effects (Fosse & Winship, 2019; Glenn, 2005; Harding, 2009; Yang, 2011; Yang & Land, 2013). These historical methods include the reduced two-factor APC model; performing nonlinear parametric transformations on at least one of the three age, period, and cohort effects; the constrained generalized linear APC model; and the age-period-cohort characteristics model.

Following the review of these methods, I next discuss their limitations and introduce the hierarchical age-period-cohort cross-classified random effects model (HAPC-CCREM). Given that the HAPC-CCREM is an advanced form of multilevel modeling, a broad overview of multilevel models is provided as context for the HAPC-CCREM. Additionally, a review of applied studies utilizing the HAPC-CREM is provided to demonstrate the utility and uptake of the model in the literature. Next, specific attention is given to methodological developments of the HAPC-CCREM. Namely, previous methodological studies and reviews have examined the following as it relates to the HAPC-CCREM: guidelines for centering variables; the accuracy of restricted maximum likelihood estimation epirical Bayes (REML-EB) estimates; assumptions of independence; and a model building approach for assessing the significance of the period and cohort effects. Despite these important advancements, some criticisms of the HAPC-CCREM exist. These criticisms are reviewed in light of responses from proponents of the HAPC-CCREM, who refute the criticims by examining the erroneous simulation studies in which they are based.

In statistical research it is common to provide abbreviations for models with long names to alleviate redundancy. Perhaps because the research on estimating age, period, and cohort effects stems from several different disciplines, naming conventions and abbreviations in the literature are often divergent, although specific researchers tend to be consistent throughout their own work. In this chapter and throughout the dissertation, I blend the abbreviations given to *model name*s utilized by Fosse and Winship (2019) and Yang and Land (2013). Readers are directed to the List of Abbreviations as a placeholder for model names. Regarding *statistical notation*, I keep with the notation utilized by developers or major proponents of the specific model under review. For example, notation for the HAPC-CCREM follows Yang and colleagues (Frenk et al., 2013; Yang, 2006, 2008; Yang & Land, 2013, 2006), while notation used in reviewing hierarchical linear models follows Raudenbush and Bryk (2002). Additionally, the terms multilevel models and hierarchical linear models are used interchangeably.

Finally, a note on the use of the term *cohort*. In many disciplines, such as public health and demography, a cohort is defined as a group of entities or individuals collectively experiencing a shared event during a specified time frame (Yang, 2011). For example, a group of individuals that acquired a disease in an outbreak, or a group of students that graduated high school in the same year, or a group of individuals that were married in a given year, are all examples of cohorts. In the area of life course research specifically concerned with age-period-cohort models, the particular cohort of interest is the birth cohort, or all individuals born during a given year. As discussed in Chapter 1, the birth cohort is specifically of interest because it indexes commonly shared experiences tied to a specific historical era that an individual carries with them throughout life (Alwin & McCammon, 2003) Therefore, throughout this dissertation the use of the term co-hort specifically implies the birth cohort.

Historical Methods of Estimating APC Models

Consider the classical APC (C-APC) model that is statistically defined as:

$$Y_{ijk} = \mu + \alpha_i + \pi_j + \gamma_k + \varepsilon_{ijk}, \tag{2}$$

where Y_{ijk} represents the outcome variable of interest for person *i* in period *j* and cohort *k*, μ is the regression intercept, or the expected outcome when all variables in the model equal zero, α_i represents the *i*th age effect, π_j represents the *j*th period effect, γ_k represents the *k*th cohort effect, and ε_{ijk} represents the error term (Fosse & Winship, 2019). Historically, the C-APC model was of most interest for use on aggregated dataset where Y_{ijk} represented specific incidence and prevalence rates for an outcome of interest, such as birth and death rates (Fosse & Winship, 2019; Yang & Land, 2006)

Perhaps the easiest way to examine age, period, and cohort effects is to estimate a reduced two-factor APC model (Yang & Land, 2013). Conceptually, this means that at any one time a researcher can examine two of the three effects in the C-APC model, and then sequentially estimate different models until all combinations of the age-period-cohort array have been examined. This results in a re-specification of the C-APC model into the following series of equations, in which results could be compared across all three resulting models:

$$\begin{gathered} Y_{ik} = \mu + \alpha_i + \gamma_k + \varepsilon_{ik} \\ Y_{jk} = \mu + \pi_j + \gamma_k + \varepsilon_{jk} \\ Y_{ij} = \mu + \alpha_i + \pi_j + \varepsilon_{ij} \end{gathered} \tag{3}$$

Although this approach is technically not a solution to the APC identification issue, it may prove useful if an outcome substantially varies in one or two of the effects over the third. As pointed out by Glenn (2002, 2005), even cross-tabulations and graphical examinations for two out of the three APC effects at a time may provide very insightful descriptive information about each of the effects.

If these types of descriptive analyses indicate that one or two of the effects can be prioritized over the third, there are further useful statistical models that align well with life course sociological theory. The most popular of these procedures has been the linear decompositions set forth by Firebaugh to examine intracohort change and cohort succession effects (Firebaugh, 1997). Specifically, for an outcome of interest, the following regression equation for the C-APC model is specified:

$$Y_{jk} = \mu + \pi_j + \gamma_k + \varepsilon_{jk}, \tag{4}$$

which provides resulting regression coefficients for the period and cohort effects. That is, the model excludes the age parameter. The values for the period and cohort effects are then used to partition longitudinal trends into intracohort change and cohort succession effects. Intracohort change is calculated as

$$\pi_j(\text{Period}_\text{T} - \text{Period}_1), \tag{5}$$

where Period_T represents the value of the final survey year and Period_1 represents the value of the first survey year. In a similar manner, cohort succession effects are calculated as

$$\gamma_k\left(\overline{\text{Cohort}_T} - \overline{\text{Cohort}_1}\right), \tag{6}$$

where $\overline{\text{Cohort}_T}$ represents the average year of birth for the last time period and $\overline{\text{Cohort}_1}$ represents the average year of birth for the first survey year. These values are then summed to represent the total change over time:

$$\Delta Y_{jk} = \pi_j \Delta \mathrm{T} + \gamma_k \Delta \mathrm{C} + \varepsilon_{jk}. \tag{7}$$

where $\pi_j \Delta \mathrm{T}$ represents the intracohort change effects derived from Equation 5, $\gamma_k \Delta \mathrm{C}$ represents the cohort succession effects derived from Equation 6, and ε_{jk} represents the error term.

In addition to the reduced two-factor APC model and linear decomposition techniques, another historical method of estimating APC effects is to conduct a nonlinear parametric transformation for at least one of the three age, period, and cohort effects in the C-APC model as described by Mason and Fienberg (1985). This is usually carried out by including quadratic or cubic terms for the age effect, but polynomial effects for the periods and cohorts can also be specified. While specifying a polynomial term for one of the three APC effects breaks the linear dependency described in the APC identification issue, the model is limited based on its strong theoretical assumption that the functional form specified mirrors the corresponding effect in real life (Winship & Harding, 2008). Therefore, the nonlinear parametric transformation approach may produce strikingly misleading or differing results depending on which of the effects are transformed, particularly if more than one of the effects is transformed at a time.

Given the limitations of the approaches discussed so far, several authors (Fosse & Winship, 2019; Glenn, 2005; Harding, 2009; Yang, 2011; Yang & Land, 2013) contend that the most widely utilized historical method for APC modeling is the constrained generalized linear APC model (CG-APC) proposed by Mason and colleagues (Fienberg & Mason, 1979; Mason et al., 1973; Mason & Fienberg, 1985). The model is generated as follows. First, each age, period, and cohort variable is separated into specified categorical widths of interest to the researcher, where the cohort and period groups are set to longer widths than the age groupings to break the linear dependency among the three terms. Next, dummy indicators are constructed for each individual to reflect their corresponding age, period, and cohort membership. For each age, period, and cohort set of dummy indicators, one grouping must be excluded from the analysis as the referent category, which is analogous to the coding of categorical variables in traditional regression modeling. Then, an equality constraint is placed on one additional age, period, or cohort grouping by leaving it out of the regression model. Typically, an adjacent category is chosen that sits next in sequence to the overall referent category for the additionally chosen equality constraint (Mason & Fienberg, 1985). This approach is theoretically substantiated on the grounds that in real life, two adjacent social groupings should theoretically be more similar than two groups that are spaced further apart. For example, the real life trends exhibited in the cohorts "1900-1905" and "1906-1910" would be more similar than imposing the equality constraint on the "1900-1905" and "2010-2015" cohorts (Glenn, 2005).

As an illustration of the CG-APC model, consider a hypothetical dataset with age categories $A_1, A_2, \dots, A_{30}$, period categories $P_1, P_2, \dots, P_{25}$, and cohort categories $C_1, C_2, \dots, C_{20}$. Using this dataset, several possible re-specifications of the classical APC model to the constrained generalized linear APC model are:

$$\begin{aligned}
Y_{ijk} &= \mu + \sum_{A_n}^{A_{28}} \alpha_i + \sum_{P_n}^{P_{24}} \pi_j + \sum_{C_n}^{C_{19}} \gamma_k + \varepsilon_{ijk}, \\
Y_{ijk} &= \mu + \sum_{A_n}^{A_{29}} \alpha_i + \sum_{P_n}^{P_{23}} \pi_j + \sum_{C_n}^{C_{19}} \gamma_k + \varepsilon_{ijk}, \\
Y_{ijk} &= \mu + \sum_{A_n}^{A_{29}} \alpha_i + \sum_{P_n}^{P_{24}} \pi_j + \sum_{C_n}^{C_{18}} \gamma_k + \varepsilon_{ijk}.
\end{aligned} \tag{8}$$

In the first re-specification, the additional equality constraint is placed within the set of dummy indicators for age, while in the second and third re-specifications, the additional equality constraint is placed within the dummy indicators for period and cohort, respectively. The allure and benefit of the CG-APC model is that it allows all effects to be estimated simultaneously while not requiring nonlinear parametric transformations of the age, period, and cohort effects. However, the challenge is that the model behaves differently depending on which set of equality constraints are imposed, and therefore strong theoretical assumptions must be made *a priori* about which constraints will most closely reflect real life trends and minimize the error produced by the model (Glenn, 2005; Yang et al., 2004). Moreover, the specification of APC effects using dummy indicators does not provide an exact estimate for a particular age, period, or cohort of interest, but instead the parameters always reflect a comparison of one category to the referent within the model.

Following the introduction of the CG-APC model in the early 1970s, one slightly more recent APC model that has proven popular over time is the age-period-cohort characteristics (APC-C) model introduced by O'Brien and colleagues (O'Brien, 1989, 2000; O'Brien et al., 1999). The premise behind the APC-C model is that proxy variables representing demographic characteristics can be substituted for the cohort effects parameters in the classical APC model. For example, instead of using the linearly derived cohorts like in the C-APC model or dummy indicators like in the CG-APC model, one could use characteristics like average educational level or proportion of unemployed individuals within each cohort. In effect, the proxy variables approach replaces the temporal cohort variable with a separate non-temporal variable that represents some aspect of the cohorts (Fosse & Winship, 2019). However, the use of proxy variables for the cohort effects is only worthwhile if multi-collinearity does not exist between the proxy variable and the age and period effects. Otherwise, the model potentially returns to the same identification issue of the C-APC model.

Collectively, these four approaches represent the main ways that researchers have examined age, period, and cohort effects for the past thirty to forty years. While useful in various ways, none of the approaches fully solve the identification issue by allowing estimates for all APC effects to be estimated at once, and each approach has its own benefits and drawbacks as discussed above. Even with the more statistically savvy methods such as the constrained generalized age-period-cohort model and the age-period-cohort characteristics model, “side information” and additional “strong theory” is usually required to supplement and complete any interpretations drawn from the analysis, given that the identification issue has not been fully resolved (Glenn, 2002, 2005; H. Smith et al., 1982). Considering the limitations of these historical methods, a tremendous breakthrough came in resolving the APC identification problem with the research of Claire Yang (formerly Yang Yang) and Kenneth Land. Yang and Land (2006) proposed that all three effects could be estimated at one time utilizing an advanced multilevel modeling technique known as cross-classified random effects modeling (CCREM). Given that CCREMs fall into the broader class of hierarchical linear models, they termed their solution the hierarchical age-period-cohort cross-classified random effects model (HAPC-CCREM). In the section that follows, important conceptual, methodological, and statistical motivations behind the HAPC-CCREM are outlined. An overview of multilevel models and their extensions is also provided as a fundamental background to the HAPC-CCREM.

The Hierarchical Age-Period-Cohort Cross-Classified Random Effects Model

Fundamental to Yang and Land’s solution to the identification problem is the estimation of age, period, and cohort effects derived from repeated cross-sectional data. Contrasted with a panel study design or prospective cohort study design, in which the same set of individuals are followed over time and surveyed at multiple time points, a repeated cross-sectional data study design samples a new population of individuals on a regular basis, each time creating a different sample of individuals that are representative of that cross-section, while often asking identical questions across cross-sections (Brady & Johnston, 2015). Yang and Land (2006) contend that the repeated cross-sectional study design presents both opportunities and challenges for age-period-cohort modeling. Specifically, the micro-level data collected at each cross-section can be aggregated to the population level to conform to earlier APC methods (for example, see Fienberg & Mason, 1979; Heckman & Robb, 1985; K. O. Mason et al., 1973) while still retaining individual-level information regarding the dependent variable of interest and associated control or predictor variables. However, the challenge lies in how to take advantage of this data structure. To that end, their proposed solution is the mixed models approach they set forth to address the APC identification issue.

Apart from the conceptual justification that repeated-cross sectional data allows the incorporation of both individual and aggregate demographic information, their use of the mixed models approach is also justified from a statistical standpoint as well. Yang and Land (2006) contend that the historical ways of overcoming the identification issue, including equality constraints, the use of proxy variables, or transforming at least one of the APC variables to break the linear dependency results in a single-level, pooled fixed effects regression model with the potential for biased parameter and standard error estimates. As an illustration, consider the following equations that represent the outcome of a theoretical dependent variable of interest for five individuals belonging to the five-year birth cohort 1960-1964 that were sampled from a repeated cross-section in the year 1990 (Yang & Land, 2006, pp. 83–84):

$$
\begin{aligned}
Y_{1,1990,1960-1964} &= \beta_0 + \beta_1(30) + \beta_2(30)^2 + \varepsilon_{1,1990,1960-1964} \\
Y_{2,1990,1960-1964} &= \beta_0 + \beta_1(31) + \beta_2(31)^2 + \varepsilon_{2,1990,1960-1964} \\
Y_{3,1990,1960-1964} &= \beta_0 + \beta_1(32) + \beta_2(32)^2 + \varepsilon_{3,1990,1960-1964} \\
Y_{4,1990,1960-1964} &= \beta_0 + \beta_1(33) + \beta_2(33)^2 + \varepsilon_{4,1990,1960-1964} \\
Y_{5,1990,1960-1964} &= \beta_0 + \beta_1(34) + \beta_2(34)^2 + \varepsilon_{5,1990,1960-1964}
\end{aligned}
\qquad (9)
$$

In these equations, β_0 represents the regression intercept for each person surveyed in 1990, while β_1 and β_2 represent the linear and quadratic age effects, respectively. As Yang and Land note

(2006, p. 84), completing the APC model for this theoretical outcome requires specification of the error terms, represented by:

$$\varepsilon_{i,1990,1960-1964} = \beta_{1990} + \gamma_{1960-1964} + e_{i,1990,1960-1964}, \text{ for } i = 1, 2, \ldots, 5. \tag{10}$$

The resulting model specified above is a single-level, fixed-effects regression model that breaks the linear dependency among the age, period, and cohort variables via the inclusion of the squared age parameter.

Importantly, the errors in this model are hypothesized as fixed, or constant, deviations from the regression intercept where β_{1990} represents the fixed time-period effect and $\gamma_{1960-1964}$ represents the fixed 1960-1964 birth-cohort effect. Statistically, modeling these effects as fixed results in one overall time period effect, where the period effect represents a constant deviation from the regression intercept from year to year, and one overall cohort effect, where the cohort effect represents a constant deviation from the regression intercept from cohort to cohort. Sharing this between-cluster common deviation from the intercept assumes that each cohort or survey year is conceptually similar to one another and not likely to change over time, instead of taking into account the potential that each cluster might have its own unique effect (Yang & Land, 2006). Moreover, the use of a fixed effects approach may produce biased results given that these models derive parameter estimates and standard errors only from within-cluster information, ignoring and not taking into account between-cluster variability (Allison, 2009). A mixed model, therefore, includes both fixed and random effects to surpass these limitations (Carson, 2013). Considering these issues, Yang and Land propose the use of a multilevel mixed models approach that takes into account the possibility of both fixed and random effects for the estimation of period and cohort effects.

An Overview of Multilevel Models and their Extensions

The use of a multilevel model is critical to the estimation of age-period-cohort effects as put forth by Yang and Land. Accordingly, an overview of two-level hierarchical linear models and their extensions is necessary before describing the multilevel approach advocated by Yang and Land. Multilevel models have long been utilized in the educational, social science, and public health arenas to analyze data collected from observations grouped by clusters. For example, in the field of education perhaps the most commonly encountered naturally occuring clustering mechanism is that of students nested in classrooms, or classrooms nested within schools. Single-level models that ignore these clusters violate the assumption in ordinary least squares regression (OLS) of independence of error terms, given that an outcome of interest is likely to be correlated with cluster membership (Snijders & Bosker, 2012; Woltman et al., 2012).

The simplest multilevel model is a two-level unconditional model. In Raudenbush and Bryk (2002) notation, the model at level one is specified as:

$$Y_{ij} = \beta_{0j} + e_{ij}, \;\; e_{ij} \sim N(0, \sigma^2), \tag{11}$$

where Y_{ij} represents the outcome for observation i nested in the level-two cluster j, β_{0j} is the regression intercept, or the average value of Y for all level-one units nested within the level-two cluster j, and e_{ij} is the level-one residual term associated with each observation i nested in the level-two cluster j, assumed normally distributed with mean 0 and variance σ^2. The level-two unconditional model specification is:

$$\beta_{0j} = \gamma_{00} + u_{oj}, \; u_{oj} \sim N(0, \tau_{00}). \tag{12}$$

The level-one intercept term β_{0j} is modeled as an outcome at level-two, where γ_{00} represents the grand mean, or the overall average value of Y across all level-one observations and level-two clusters, and u_{oj} represents the level-two residuals, or the random effect of each level-two cluster

j. Importantly, the random effect is assumed to be normally distributed with mean 0 and variance τ_{00}.

The unconditional multilevel model can be extended to a conditional model by including variables of interest at level-one, such as:

$$Y_{ij} = \beta_{0j} + \beta_1 X_{ij} + e_{ij},\ e_{ij} \sim N(0, \sigma^2), \tag{13}$$

and at level-two as:

$$\begin{cases} \beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{oj} \\ \beta_{1j} = \gamma_{10} + \gamma_{11} W_j + u_{1j} \end{cases}, \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix} \right)\ . \tag{14}$$

In the conditional model, Y_{ij} represents the outcome for observation i nested in the level-two cluster j. The u_{oj} term represents the unique increment to the intercept associated with cluster j, conditional upon the variables included in the model and the u_{1j} term represents the unique increment to the slope associated with cluster j, also conditional upon the variables included in the model. The level-one residual term e_{ij} represents the difference in the observed outcome for observation i and the predicted outcome in cluster j, conditional upon the variables included in the model. The additional parameters can be described as follows. The β_{0j} term represents the level-one intercept, which is modeled as an outcome at level-two. The γ_{00} term is the regression intercept, or the expected value of Y for all level-one units when the values of X_{ij} and W_j equal zero. The β_{1j} term represents the coefficient for the level-one predictor variable, which is modeled as randomly varying across level-two units through the u_{1j} parameter. The γ_{10} term therefore represents the average regression slope across the level-two units holding constant W_j. The γ_{01} term is the regression coefficient for the level-two variable that is regressed on the level-one intercept and the γ_{11} term represents a cross-level interaction between the level-one and level-two predictor variables. Additionally, the dispersion of the level-two random effects is assumed multivariate normally distributed with means of 0 and variances specified in the variance-covariance matrix $\mathbf{T}$, where τ_{00} represents the conditional variance in the level-one intercepts, τ_{11} represents the conditional variance in the level-one slopes, and τ_{01} represents the covariance between the level-two intercepts and slopes. Note that the provided level-two specification represents all theoretical modeling possibilities for the level-one parameters (randomly varying level-one intercept, randomly varying level-one slope, and a cross-level interaction for the level-two predictor that is regressed on both the level-one intercept and slope).

The two-level model described above has been extended to a variety of other relevant research designs. One particular extension of the two-level model as it relates to the mixed models approach for estimating age-period-cohort effects is the cross-classified random effects model (CCREM) proposed by Raudenbush (1993). The two-level model described above is predicated on the assumption of a pure nesting structure such as students nested within classrooms, or students nested within classrooms nested within schools. However, this assumption is not always tenable, nor have researchers only been interested in examining the effects of variables that fall within purely nested structures such as the ones described above (Goldstein, 1994). Cross-classified models allow for the analysis of impurely clustered data structures by taking into account that observations can be nested within higher order levels, but these levels are not necessarily nested within one another (Beretvas, 2011).

Following Raudenbush and Bryk (2002) notation, an unconditional CCREM can be specified at level one as:

$$Y_{ijk} = \pi_{0jk} + e_{ijk},\ e_{ijk} \sim N(0, \sigma^2), \tag{15}$$

where Y_{ijk} represents the outcome for observation i cross classified by clusters j and k, π_{0jk} is the average value of the outcome for observations nested in cluster jk, and e_{ijk} is the level-one residual, assumed normally distributed with mean 0 and variance σ^2. The level-two unconditional CCREM is represented as:

$$\begin{gathered}\pi_{0jk} = \theta_0 + b_{00j} + c_{00k} + d_{0jk}, \\ b_{00j} \sim N(0, \tau_{b00}), \\ c_{00k} \sim N(0, \tau_{c00}), \\ d_{0jk} \sim N(0, \tau_{d00}),\end{gathered} \tag{16}$$

where π_{0jk} represents the level-one intercept model as an outcome at level-two that randomly varies across clusters *j* and *k* and θ_0 represents the grand mean of the outcome across all observations. The b_{00j} parameter is the random main effect for cluster *j*, averaged over all *k* clusters, assumed normally distributed with mean 0 and variance τ_{b00}. Similarly, c_{00k} represents the random main effect for cluster *k*, averaged over all *j* clusters, assumed normally distributed with mean 0 and variance τ_{c00}.

In the unconditional CCREM, d_{0jk} is a random interaction effect, calculated as the deviation of a cell mean from the predicted grand mean and the two main effects b_{00j} and c_{00k}, assumed normally distributed with mean 0 and variance τ_{d00}. Conceptually, d_{0jk} can be thought of as a moderation effect for the clustering variables. In practice this random interaction is almost never estimated (that is, the effect is set to 0) given small within-cell sample sizes that make it difficult to partition variance between the level-one within-cell variance component σ^2 and the level-two between-cell variance component τ_{d00} (Beretvas, 2008; Rasbash & Goldstein, 1994; Raudenbush & Bryk, 2002, p. 378). Although simulation studies indicate that excluding this effect can bias random effect estimates at level-two (Lee & Hong, 2019; Y. Shi et al., 2010), some argue that the assumption of simple additive random effect parameters without the inclusion of a multiplicative effect suffices in most applied research studies (Fielding & Goldstein, 2006, pp. 29–30).

Similar to the conditional two-level model, a conditional CCREM can readily be implemented, but a "general" set of equations is not plausible given the multitude of possible effects to be modeled. That is, the introduction of two level-two clusters, instead of one purely structed clustering mechanism, results in many possible combinations for specifying fixed and random effects. Generally speaking, the following options apply. At level-one of the CCREM, independent variables of interest can be modeled in the conventional way, but decisions must now be made regarding if these effects will be fixed across both cross-classified level-two clusters, randomly varying at only one level-two cluster, or randomly varying across both level-two clusters. Independent variables at either or both of the level-two clusters may also be specified, and these predictors can be regressed on any combination of the level-one intercept or slopes. Considering the multitude of parameters that can be included in the CCREM, *a priori* theoretical and statistical consideration should be exercised with the entering of both level-one and level-two variables in the model. This holds especially true with the inclusion of random effects, where higher numbers of random effects result in more complex estimation procedures that may result in model convergence issues (Beretvas, 2011).

Specifying APC Effects as CCREM

Having provided an overview of two-level models and their extension to cross-classified random effects models, I now return to the utilization of mixed models in the estimation of age, period, and cohort effects. Yang and Land (2006) contend that data collected from a repeated cross-sectional study design, described above, results in a naturally occurring cross-classified data structure. Specifically, observations from individuals can be nested into clusters aligned with the year in which they were surveyed (period effects), which can then be cross-classified into specific birth-cohorts (cohort effects) of interest for analysis. Specifically, the age effects are modeled at level-one with linear and quadratic regression coefficients, and the period and cohort effects are derived through the random effects estimates of the cross-classified clusters. Yang and Land termed their model the hierarchical age-period-cohort cross-classified random effects model (HAPC-CCREM). Using their notation (2006), the model can statistically be represented at level one as:

$$Y_{ijk} = \beta_{0jk} + \beta_1 \mathrm{Age}_{ijk} + \beta_2 \mathrm{Age}_{ijk}^2 + \sum_{p}^{P} \beta_p X_{ijk} + e_{ijk}, \quad e_{ijk} \sim N(0, \sigma^2) \tag{17}$$

where Y_{ijk} represents the outcome for observation i cross-classified by cohort j and period k, β_{0jk} is the mean of the outcome for an individual belonging to cohort j and period k, β_1 and β_2 represent the linear and quadratic age effects, β_p is a vector of regression coefficients of the other individual-level demographic control variables X_{ijk}, and e_{ijk} represents the level-one residual, assumed normally distributed with mean 0 and variance σ^2.

The level-two specification of the CCREM applied to the estimation of age, period, and cohort effects is:

$$\begin{aligned} \beta_{0jk} &= \gamma_0 + u_{0j} + v_{0k}, \\ u_{0j} &\sim N(0, \tau_u), \\ v_{0k} &\sim N(0, \tau_v), \end{aligned} \tag{18}$$

where β_{0jk} is the level-one intercept modeled as an outcome at level-two and γ_0 is the grand mean of the outcome across all individuals. The parameter u_{0j} is the random effect parameter for the birth-cohorts, which conceptually is the contribution of cohort j to the intercept averaged over all periods, assumed normally distributed with mean 0 and variance τ_u. The parameter v_{0k} is the random effect parameter for the survey years / time periods, which conceptually is the contribution of period k to the intercept averaged over all birth-cohorts, assumed normally distributed with mean 0 and variance τ_v.

The model put forth by Yang and Land (2006) is identical to the CCREM articulated by Raudenbush (1993), except that it excludes a multiplicative interaction effect between the period and cohort random effects (i.e., a w_{0jk} term). Importantly, the HAPC-CCREM breaks the linear dependency of the identification issue by including nonlinear parametric transformations for the age effects. Additionally, the cohort and period effects are coded as class memberships at level two, where the researcher must treat the effects as categorical variables into which individuals can be grouped. Estimating APC effects in the multilevel modeling framework differs from the CG-APC approach in that dummy indicators are not utilized to represent the period and cohort effects, but instead an estimate is generated for each period and cohort that does not involve comparison to a referent period or cohort category left out of the model as in the CG-APC.

Applications of the HAPC-CCREM

Reflective of the utility found in the HAPC-CCREM, more than fifty studies have been published that utilize the method since its implementation by Yang and Land in 2006 (see Appendix A for a compendium of these studies). Researchers across diverse disciplines within the social sciences have taken advantage of the HAPC-CCREM to examine age, period, and cohort trends in topics such as attitudes toward controversial social issues, religious affiliation, gender ideology and sexual attitudes, and changes in health across the life course. Collectively, the studies have utilized a substantial amount of historical data, with many of the analyses spanning more than thirty years of historical life course trends. Almost all the studies have employed data from the General Social Survey, a popular repeated cross sections dataset used frequently in the discipline of sociology, although a few studies have used data from Monitoring the Future and the National Health Interview Survey.

One of the most frequent applications of the HAPC-CCREM has been to examine longitudinal changes in attitudes toward controversial topics attributable to age, period, and cohort effects. For example, in my own previous research (Attell, 2020), I examined how attitudes toward euthanasia and suicide for terminally ill persons changed between 1977 and 2016. Other important issues that have been examined in this area include attendance at controversial protests and political events and participation in political activities like voting (Caren et al., 2011; Fullerton & Stern, 2010, 2013; Horowitz, 2015), changing attitudes towards women's work (Donnelly et al., 2016), changing political tolerance (Eisenstein et al., 2017; Schwadel & Garneau, 2014), trends in public approval for legalizing marijuana (Schwadel & Ellison, 2017), and changing attitudes toward immigration and immigrants (Gorodzeisky & Semyonov, 2018; Wilkes & Corrigall-Brown, 2011), to name a few. A substantial portion of this research has specifically focused on issues related to gender, gender ideology, and sexuality (Donnelly et al., 2016; Pampel, 2011, 2016; Shu & Meagher, 2018).

A second major area of research using the HAPC-CCREM has been in the area of physical health, mental health, and wellbeing. In two separate studies Masters and colleagues analyzed trends in all-cause mortality attributable to gender, race, and education in conjunction with age, period, and cohort effects (Masters, 2012; Masters et al., 2012). To date, four studies have focused on explaining trends in happiness as a proxy for overall wellbeing and mental health (Bardo et al., 2017; Bardo, 2017; Twenge, Sherman, & Lyubomirsky, 2016; Yang, 2008), and one study examined changes in self-rated health (Zhang, 2017). Additionally, more narrowly focused health related outcomes such as prevalence of sexual activity (Twenge et al., 2017), marijuana use (Keyes et al., 2011), alcohol use (Keyes et al., 2012), and obesity (Reither et al., 2009) have also been examined using the HAPC-CCREM.

Beyond the areas of attitudes toward controversial social topics and changes in physical health, mental health, and wellbeing, a significant area of inquiry has also been examining perceptions of government spending and confidence in social institutions (e.g., Fullerton & Dixon, 2010). Additional HAPC-CCREM studies have shown consistent and steady period-based declines in confidence in a variety of social institutions, including media, medicine, Congress, and the Supreme Court (Twenge et al., 2017; Zheng, 2015), where some of these studies have focused on narrowing in on group differences in these trends over time (Gauchat, 2012). Somewhat related to the area of social institutions, several studies have applied the HAPC-CCREM to examine various issues related to religion (Schwadel, 2010a, 2010b; Twenge, Sherman, Exline, et al., 2016). Additionally, some attention in the literature has been devoted to examining changing attitudes regarding religion in the public sphere (Schwadel, 2013a).

Methodological Developments of the HAPC-CCREM

Having provided an overview of age-period-cohort models including predecessors to the HAPC-CCREM, Yang and Land's solution to the identification problem that has been informed by life course sociological theory, and the applied literature that has utilized the HAPC-CCREM, I now consider specific methodological advances that have been made in the development and subsequent refinement of the HAPC-CCREM. Many of these advances have been applied and conceptual discussions aimed at specific issues that arise when utilizing the model in real life, such as how to center variables in the model (Yang & Land, 2006) or how to assess the significance of the period and cohort effects (Frenk et al., 2013). Additionally, a few simulation studies examining the accuracy of the HAPC-CCREM have been conducted. However, these simulation studies sparked considerable debate and were met with great skepticism among proponents of the HAPC-CCREM, who pointed out serious flaws in the design of the simulation studies critiquing the HAPC-CCREM.

Centering Decisions in the HAPC-CCREM

While the main purpose of Yang and Land's 2006 article was to propose and articulate the use of mixed models in the estimation of age, period, and cohort effects, they afford a brief amount of space to the methodological issue of centering in the HAPC-CCREM. The issue of centering is very important to the interpretation of all types of multilevel models and has been given considerable attention in the broader methodological literature (Enders & Tofighi, 2007; Hofmann & Gavin, 1998; Paccagnella, 2006). Just as the case with a single-level ordinary least squares (OLS) regression equation, in multilevel models the level-one intercept coefficient (usually specified as randomly varying across level-two units) can be interpreted as the average outcome for an individual when all predictor variables equal zero (Raudenbush & Bryk, 2002). Accordingly, when predictor variables do not contain a zero value in the actual data, these parameters must be rescaled, or "centered", so that the intercept can be appropriately interpreted.

The decision on which type of centering to use is not always straightforward, and depends on both theoretical and statistical considerations (Enders & Tofighi, 2007). In the case of utilizing CCREM to estimate age, period, and cohort effects, we must consider if level-one predictor variables will utilize their natural metric, be group-mean centered at the cohort or period clusters, or be grand-mean centered for the entire sample. In proposing their model, Yang and Land (2006) provide an applied example of the CCREM for analyzing age, period, and cohort trends in verbal ability in the United States from 1974-2000, controlling for education, gender, and race. They contend that previous research demonstrates "pure" age effects in verbal ability that do not depend on cohort membership, and therefore grand-mean center the age effects in their model. However, they cite previous research indicating between cohort differences in educational attainment over time, such that newer birth cohorts achieve higher levels of education, thereby justifying their use of group-mean centering the education variable across cohort clusters. In the case of gender and race, dummy indicators are utilized to contrast differences in verbal ability between male (referent) and female respondents, and white (referent) and black respondents. Because a typical 0 / 1 coding scheme for the indicators was utilized, zero values were present in the variables and centering was not necessarily required. Note, however, that dummy indicators may still be group-mean or grand-mean centered, and these centering decisions influence the interpretation of the level-one intercept and its associated period and cohort variance components (see Enders & Tofighi, 2007, pp. 134–135). This discussion of centering in the HAPC-CCREM indicates that previous research and theory must be drawn upon, in addition to examining the actual values in the data, when making decisions about how to center.

Accuracy of REML-EB Estimates in the HAPC-CCREM

In the HAPC-CCREM, estimates of the fixed effects parameters for the age effects (and other level-one explanatory variables) are obtained by restricted maximum likelihood estimation, and the period and cohort effects are obtained by empirical Bayes estimates of the level-one fixed effects. Taken together, these procedures comprise the "REML-EB" estimates for age, period, and cohort effects. For multilevel modeling in general, REML-EB estimation is appropriate when there are a large number of level-two clusters that contain relatively equal sample sizes across clusters (Raudenbush & Bryk, 2002, pp. 410–411). In the HAPC-CCREM, however, existing datasets may not contain enough survey years and birth cohorts to meet the large number of clusters required by REML-EB estimation procedures to produce accurate estimates in the model. To address this issue, Yang (2006) conducted a simulation study to examine if the accuracy of the REML-EB estimates produced by the HAPC-CCREM depend on the number of survey years available for analysis.

The simulation study was informed by a real data analysis of trends in verbal intelligence (the “WORDSUM” test) in the United States between 1974 and 2000 based on fifteen repeated cross-sections of the General Social Survey. The WORDSUM test was used in a simple additive manner as the dependent variable by summing across the ten items in the test. In the model, verbal intelligence was regressed on linear and quadratic age terms that control for the effect of education, sex, and race. At level-two, the proportion of individuals in each five-year birth cohort who read newspapers on a daily basis and the average hours of TV watching per day within each five-year birth cohort were included as predictors. All predictor variables were treated as fixed effects in the model. The REML-EB parameter estimates from the real data study were then used in the data generating models for the simulation.

One experimental condition with two levels was specified. For the first, “total sample” level of the experimental condition, the data generating model was used to generate 1000 datasets with 19 five-year birth cohorts and 15 single-year periods, as this represented the cohort and period groupings in the real data. For the second, “thinned sample” level of the experimental condition, the data generating model was used to generate 1000 datasets with 5 five-year birth cohorts and 5 single-year periods. The rationale behind the second experimental condition was to examine the accuracy of the REML-EB estimates for datasets with a limited number of birth cohorts and survey years, that is, a smaller dataset.

Several outcomes for the simulation study were of interest (Yang, 2006). In both the total sample and thinned sample, the mean fixed effects parameters from the data estimating models were almost identical to the true parameter values, lending preliminary support that the REML-EB fixed effects estimates were accurate in large and small datasets. In the total sample, the means of the individual, period, and cohort random effects’ variance component estimates were almost identical to their corresponding true variance component estimates. However, in the thinned sample, the mean of the individual random effect variance component was slightly smaller than the true value, and the means of the period and cohort random effects' variance components were slightly larger than the true values.

In addition to these qualitative, "eye-ball" comparisons, mean square error (MSE) was also examined. In the total sample, MSE was relatively minimal among the fixed effects and the cohort, period, and individual variance component estimates. However, in the thinned sample, MSE was large for the intercept, News, TV, and age parameter estimates and small for the remaining fixed effects estimates and cohort, period, and individual variance component estimates. For the fixed effects in the thinned sample where MSE was high, the magnitude of difference in their corresponding result from the total sample was substantial. For example, in the total sample the MSE for the TV parameter was 0.0622, while MSE for the TV parameter in the thinned sample was 3.9250. For each parameter estimated in the simulated datasets, coverage rates were also examined by calculating the proportion of 95 percent confidence intervals for each parameter that included the corresponding true value. Coverage rates for the coefficient estimates were close to 95 percent in the total sample but slightly lower than the nominal level in the thinned sample, ranging from 91% to 95%. Coverage rates were close to the nominal level for the individual, period, and cohort variance components in the total sample, ranging from 95% to 96%, but the coverage rates for the cohort and period variance components were much lower than 95 percent in the thinned sample (65% and 44%, respectively).

Mean estimates, MSE, and coverage rates were also examined for the random effects estimates associated with each period and cohort cluster in the the total (19 cohorts and 15 periods) and thinned (5 cohorts and 5 periods) samples. Regarding mean estimates for the period and cohort random effects, almost all of the mean values in both the total and thinned samples greatly differed from the true values obtained from the real data study. The MSEs were also very large for each period and cohort random effect estimate across the total and thinned sample. Coverage rates associated with the random effects estimates in the total sample ranged from 40 to 60 percent for the cohort effects and 50 to 60 percent for the period effects. Coverage rates associated with the random effects estimates in the thinned sample ranged from 48 percent to 73 percent for the cohort effects and 54 to 57% for the period effects. Taken together, Yang (2006) concluded that the results of the simulation indicated that the HAPC-CCREM was not appropriate for datasets with a small number of birth cohorts and time periods. However, for larger datasets, the HAPC-CCREM produces accurate estimates for the fixed effects regression parameters and the period and cohort variance components, but the random effects parameters associated with each period and cohort cluster were not as accurate.

Assumptions of Independence for the HAPC-CCREM

After examining the accuracy of REML-EB estimates given the amount of data available for analysis, some of the more nuanced assumptions behind the model were examined. The first of these assumptions refers to the relationship between the level-one predictor variables and the level-two random period and cohort effects. In multilevel models, any level-two random effect is assumed to be independent of the set of level-one predictors (Grilli & Rampichini, 2015, 2018). Therefore, as it relates to the HAPC-CCREM, it is necessary to examine if this assumption of independence holds true. The independence assumption can be examined in one of two ways. One option is to visually inspect the relationship between the level-two random effects and each level-one predictor using a residual scatter plot and a corresponding Pearson correlation coefficient or smoothed spline to quantify the relationship (Snijders & Berkhof, 2008). If any extreme evidence of a linear or non-linear relationship exists, the independence assumption is violated, the random effect parameters should not be estimated, and a fixed effects specification should be considered (Ebbes et al., 2004)

A second, and perhaps more robust, statistical method of assessing the independence assumption is the Hausman test (Ebbes et al., 2004; Fielding, 2004). The Hausman test is a Wald chi-square test that can be used to compare a competing fixed and random effect model. As it relates to the HAPC-CCREM, the null hypothesis for the Hausman test is that the period and cohort random effects in the CCREM are independent of the level-one predictor variables. Performing this test requires the specification of the age-period-cohort model as both a random and fixed effects model with an identical set of predictor variables between the two. Yang and Land (2008) used a HAPC-CCREM and a hierarchical age-period cohort cross-classified fixed effects model (HAPC-CCFEM) with the same set of level-one predictor variables present. However, the level-two specifications are quite different. In the HAPC-CCREM the cohort and period effects are estimated as random effects that are captured by u_{0j} and v_{0k}, whereas in the HAPC-CCFEM, a set of dummy indicators representing membership in five-year birth cohorts and one-year period groups are specified at level-two. To avoid a fully saturated model, the last five-year cohort cluster and one-year period cluster were excluded as the referent groups of comparison. Regarding the Hausman test, they failed to reject the null hypothesis of no systematic differences in the parameters (Yang & Land, 2008, p. 318). Accordingly, they contend in their study that the random effects u_{0j} and v_{0k} are uncorrelated with the level-one predictors and that the HAPC-CCREM is a preferred modeling procedure to the HAPC-CCFEM given this finding.

Assessing the Significance of Period and Cohort Effects in the HAPC-CCREM

The most recent methodological development provided by proponents of the HAPC-CCREM was a general framework for assessing the significance of the period and cohort effects derived from the model (Frenk et al., 2013), which is helpful for applied researchers. Once the

HAPC-CCREM is estimated, we can examine the estimates of $X = J + K$ random effects, where J represents the number of level-two cohort clusters and K represents the number of level-two period clusters. Each random effect estimate represents the respective cohort or period deviation from the intercept. Given the number of estimates produced by the model, a framework for assessing each effect in isolation, as well the overall trend for the estimates is necessary. To address these issues, Frenk et al. (2013) proposed a four step process, which is briefly described.

First, an examination of any patterns or trends in the period and cohort effects is done by plotting each random effect estimate with an associated 95% confidence interval. This is accomplished by extracting each random effect estimate from the model, calculating a lower and upper bound for the confidence interval, and adding to the results the constant term from the estimated intercept. Second, the statistical significance of each cohort and period random effect estimate is conducted, where failing to reject the null hypothesis signifies a cohort or period effect that is not different from the overall average. Third, the deviance value and variance components are examined. Note that the analysis of the deviance value is more so related to the statistical significance of the overall model. On the other hand, the decomposition of variance components points toward practical significance of the period and cohort effects by quantifying the proportion of total variability attributable to each effect. Fourth, a global F-test is conducted on the variance components to assess the significance of the period and random effects' variance components. Failure to reject the null hypothesis would indicate a situation in which the birth-cohort or time-period effects' variance components do not significantly differ from zero. However, because the F-test only takes into account one variance component at a time, it should be performed with caution. Significance tests such as the likelihood ratio test, which takes into account all parameters in the model, would also a worthwhile analysis (Raudenbush & Bryk, 2002, pp. 63–65). Combining steps one through four provide a great deal of insight regarding the results of the HAPC-CCREM and provide a compelling framework from which to assess the period and cohort effects derived from the model.

Criticisms and Responses

Innovation does not come without critique. In a series of papers, social scientists Bell and Jones (Bell & Jones, 2013, 2014a, 2014b, 2015) utilized various simulations to call into question the findings produced by studies that utilized the HAPC-CCREM. Broadly, their simulations were designed to address two main areas of concern. The first area of concern is related to how cohorts are treated as categorical variables that define cluster membership at level-two in the HAPC-CCREM. Bell and Jones contend that treating cohorts as categorical, cluster defining variables represents a modeling choice that may not capture the true cohort effects that exist in the data (2014a). In Yang's (2006) simulation discussed previously, datasets were generated using five-year birth cohorts, to which estimating models with five-year birth cohorts were also applied. Bell and Jones (2014a) were therefore motivated to examine the accuracy of estimates from the HAPC-CCREM when cohort groupings in the estimating model did not match cohort groupings used in the data generating model. The second area of concern is related to the model's ability to detect period and cohort trends. Recall that the period and cohort effects in the HAPC-CCREM are estimated as random effects that are assumed to be normally distributed deviations from the model intercept. Given this normality assumption, Bell and Jones (2014a) tested the model's ability to recover linear period and cohort trends when they were specified in the data generating model, an issue previously unexamined by Yang (2006).

In one study, the primary interest was the accuracy of the parameter estimates and the values of the cohort and period random effects under several experimental conditions (Bell & Jones, 2014a). First, they examined the extent of biased estimates if an HAPC-CCREM was estimated on a generated dataset where no linear period and cohort trends were present (scenario 1). Second, they examined accuracy of the estimates when a linear cohort trend existed in the generated datasets (scenario 2). Third, they examined accuracy when a linear cohort trend with 1-year cohort groups was specified in the generated data, but five-year cohorts were specified in the HAPC-CCREM fitted to the simulated dataset (scenario 3). They next examined the situations in which five and seven-year cohorts were present in the simulated dataset, but only five-year cohorts were specified in the HAPC-CCREM fitted to the simulated datasets (scenarios 4 and 5, respectively). They then examined the scenario in which a linear cohort trend was specified in the simulated datasets and one-year cohort groups were specified as random effects and fixed effects in the HAPC-CCREM applied to the simulated datasets (scenario 6). Finally, they examined the impact of bias on the parameter estimates when a linear period trend was specified in the data generating model but no period effects were specified in the HAPC-CCREM (scenario 7). For each simulation condition, 1,000 randomly generated datasets were created, each of which contained 20,000 respondents with randomly distributed ages between 20 and 60 and randomly distributed periods between 1990 and 2010. In each dataset, cohorts were then calculated based on each respondents' randomly assigned age and period values.

The results of their simulations were as follows. For scenario 1, where no period and cohort trends existed in the data generating process, the HAPC-CCREM performed well with no exhibited biases in the model parameters. In scenario 2, where a linear cohort trend was specified in the data generating models, the parameters for cohorts and linear age effect were biased downwards, and the model produced an entirely linear period trend that was not present in the data generating process. In scenario 3, where a linear cohort trend existed in the data generating model and five-year cohorts were specified in the HAPC-CCREM, the linear age effects and cohort terms were once again negatively biased. In scenario 4, where five-year cohorts were specified in both the data generating and estimating models, parameter estimates were unbiased. When the cohorts were mismatched in scenario 5, the same biased coefficients were found from scenarios 2 and 3. In scenario 6, when fixed and random effects were specified for the cohort effects in the estimating models, the bias in the age effect from scenarios 2 and 3 was resolved. Finally, in scenario 7, where a linear period trend was specified in the generating model but no fixed period effects were specified in the HAPC-CCREM, they found that the period trend in the simulated dataset was redistributed between the age and cohort effects, which were overestimated. Based on these findings, Bell and Jones concluded that the HAPC-CCREM performed poorly. Notably, the model failed to detect linear period or cohort trends from the data generating models, and only produced unbiased estimates when the cohort groupings used in the estimating models matched the cohort groupings from the data generating models.

In a second simulation study, Bell and Jones (2014b) used Reither et al.'s (2009) HAPC-CCREM study related to trends in the prevalence of obesity to continue examining the issue of cohort groupings and recovery of the age, period, and cohort trends in the HAPC-CCREM. Similar to their previous study, they manipulated the cohort groupings in the data generating models and the HAPC-CCREMs applied to the simulated data. They examined three grouping scenarios. In scenario 1, no grouping was conducted in the data generating model or the fitted HAPC-CCREM. In scenario 2, no grouping was conducted in the data generating model but 5-year cohorts were specified in the HAPC-CCREM estimating model. In scenario 3, 7-year birth cohorts were specified in the data generating model but 5-year birth cohorts were used in the HAPC-CCREM estimating model. Across all scenarios, a linear trend was specified for the cohort effects in the data generating models. For each scenario, 1,000 randomly generated datasets were created.

Across all scenarios, the outcome of interest was whether the HAPC-CCREM fit to the simulated data produced the same age, period, and cohort trends present in the simulated data. The results of the simulation study indicated that across all scenarios the HAPC-CCREM fit to the simulated data produced consistently lower probabilities of obesity based on age, that the real cohort trend was not detected, and that an erroneous linear period effect was produced that did not exist in the "true" data from the data generating models. For these reasons, Bell and Jones cautioned researchers against utilizing the HAPC-CCREM.

Proponents of the HAPC-CCREM responded to these simulations with great skepticism. In a separate study, Reither and colleagues (2015) conducted their own simulations to counter those conducted by Bell and Jones. They hypothesized that Bell and Jones' simulations produced erroneous and misleading results because the period and cohort trends specified in their data generating models were *perfectly* linear, a characteristic that does not reflect real world data and trends. Therefore, they specified new data generating models that did not rely on perfectly linear period and cohort trends. Their counter simulation consisted of three such scenarios. In one scenario, seven-year cohorts were specified in the data generating models with a near linear cohort trend, and five-year cohorts were specified in the estimating models. In another scenario, three-year cohorts were specified in the data generating models with a near linear cohort trend, while five-year cohorts were specified in the estimating models. Finally, the last scenario utilized seven-year cohorts in the data generating models with a near linear period trend, while five-year cohorts were specified in the estimating models. In their counter simulation study, they found that across all scenarios the HAPC-CCREM estimating models did in fact recover the true APC effects from the data generating models, and that moreover the cohort groupings in the estimating model did not have to match the cohort groupings in the generating models to accurately recover the true APC effects (Reither et al., 2015). These findings highlight the importance of using values reflective of real data and trends in simulation studies, and lead Reither and colleagues (2015) to conclude that the HAPC-CCREM is a robust model that can, in fact, detect age, period, and cohort trends, even if we do not know the true cohort grouping that exists in the data used by the model.

The Current Study

Taken together, the previous methodological research on the HAPC-CCREM has provided meaningful insights to the model. The more demonstrative methodological research has aimed to provide guidelines for centering variables in the model (Yang & Land, 2006), assessing and visualizing the significance of the period and cohort effects (Frenk et al., 2013), and examining assumptions of independence between the level-one predictor variables and the level-two random period and cohort effects (Yang & Land, 2008). Simulation studies for the HAPC-CCREM have focused on the accuracy of estimates from the model given the size of the dataset (Yang, 2006), and if the model can recover the age, period, and cohort trends in various scenarios (Bell & Jones, 2014a, 2014b; Reither et al., 2015). However, important methodological questions remain that have not been examined but could potentially be very useful to applied researchers.

One such issue is related to the accuracy of the model parameters given the amount of data available for analysis. Recall that using the real data study of trends in verbal ability in the U.S., Yang (2006) simulated and applied the HAPC-CCREM to 19 five-year birth cohorts and 15 single period survey years, derived from repeated cross-sectional data from the General Social Survey between 1974 and 2000. These were then compared to a sample of 5 five-year birth cohorts and 5 single-year period effects, comprising a “thinned sample”. Broadly, the results of the simulation indicated that for datasets comprised of 15 repeated cross-sections, REML-EB estimates from the HAPC-CCREM were accurate in estimating the individual and cohort fixed effects regression parameters and the individual, period, and cohort variance components. For datasets comprised of 5 repeated cross-sections, REML-EB estimates from HAPC-CCREM were accurate in estimating most of the individual and cohort fixed effects regression parameters. However, in the smaller datasets the level-1 age effect, as well as the period and cohort variance components were overestimated, while the individual variance component was underestimated. Importantly, beyond Yang’s simulation, none of the other methodological studies examined the impact that the amount of survey years available for analysis may have on the accuracy of the model estimates.

There are at least three reasons why re-examining how the number of repeated cross-sections may influence the accuracy of model estimates is important. First, since Yang’s simulation was published, the data collected by the General Social Survey and other repeated cross sectional datasets have grown, which is reflected in more recently published analyses. For example, Anderson et. al’s (2017) analysis of changing attitudes towards the death penalty utilized 27 repeated cross-sections from the GSS, and Carlisle and Clark’s (2018) analysis of the associations between religion and environmentalism utilized 29 repeated cross-sections from the GSS. As publicly available datasets like the GSS continue to grow, what constitutes a “large” dataset will continue to evolve with the availability of more repeated cross-sections. Given that certain analyses using GSS data have, in fact, almost doubled the number of repeated cross-sections used in Yang’s (2006) simulation study, it is worthy of investigation to revisit and expand Yang’s study to examine the accuracy of model estimates given larger datasets.

Second, the findings of Yang's study imply that applied researchers should only specify the HAPC-CCREM on datasets with at least 15 repeated cross-sections. It may be the case that a dataset with the number of repeated cross-sections between five and 15 may also accurately recover model parameters, but currently this has not been examined in the methodological literature. Third, Yang's simulation did not include a level-2 period predictor variable, instead only focusing on a level-2 cohort predictor variable. In the applied literature, the use of period-level predictor variables may be of interest. For example, Anderson et. al's analysis of attitudes towards the death penalty included period-level predictors for the annual violent crime rate and unemployment rate (Anderson et al., 2017) and Johson and Schwadel's (2019) analysis of support for environmental spending included period-level predictors for the number of New York Times articles on the enviroment and years in which a democrat was president. The accuracy of coefficient estimates of period-level predictors such as these may be directly impacted by the number of survey years included in the analysis, but to date this issue has not been explored. Considering these three issues, the first aim of this dissertation is to answer the following question:

RQ1: What is the accuracy of REML-EB estimates at all levels in the HAPC-CCREM given the number of survey years available for analysis?

One of the interesting methodological decisions required in the HAPC-CCREM is the selection of cohorts to be used in the model, hereafter referred to as the *cohort selection mechanism*. Recall that previous research focused on the ability of the HAPC-CCREM to recover age, period, and cohort trends when the cohort selection mechanism used in the estimating model did not match the data generating model. The findings of Reither et al. (2015) indicated that the HAPC-CCREM could detect age, period, and cohort trends even if the cohort selection mechanism between the estimating model and data generating model did not match. Considering that the cohort selection mechanisms do not have to match, questions remain about which cohort selection mechanism should be used, and if one cohort selection mechanism produces more accurate estimates than the others.

In the applied literature that utilizes the HAPC-CCREM, the most commonly utilized cohort selection mechanisms are the three-, five-, or ten-year birth cohorts. To date, these decisions have been arbitrary and appear to be left to personal preference, when the decision of which cohort selection mechanism to uptake is even described or justified. Indeed, 27 out of the 57 applied studies using the HAPC-CCREM, or just 47%, provide justification for the cohort selection mechanism. The most common explanation, when one exists, references Yang's study of APC-based social inequalities in happiness in the United States, where she claimed "we can use single years of age, time periods corresponding to years in which the surveys are conducted, and birth cohorts defined by five-year intervals, which are conventional in demography" (Yang, 2008, p. 210). Beyond "convention", some researchers contend that the cohort selection mechanism should be based on substantial theoretical interests. For example, two separate studies (Kowske et al., 2010; Shu & Meagher, 2018) drew on generational theory to drive their cohort selection mechanism, abandoning the five-year birth cohort approach for much wider generational cohorts that grouped individuals into cohorts such as the GI Generation (those born between 1901-1924), Baby Boomers (those born between 1943-1960), Gen Xers (those born between 1961-1981), and Millenials (those born between 1982-2003).

The last justification for the cohort selection mechanisms used in the applied literature has included reference to statistical consideration of sample sizes within cross-classified cells. Consider the following justifications for the cohort selection mechanism:

> Data collected over time can be analyzed in many ways, including grouping by 20-year generation blocks, by decades, or by individual year. We believe that separating the data into 5-year intervals provides the best compromise between specificity and breadth. (Donnelly et al., 2016, p. 45)

> We grouped people by birth decade as a compromise between breadth and depth. Using a larger span (for example, a 20-year generation) risks losing discriminatory power, and a smaller span (such a 5-year groups) risks low sample size. (Twenge et al., 2017, p. 435)

> As is typical in cohort analyses, five-year birth cohorts are used, providing a range that is wide enough to provide reliable statistical estimates but narrow enough to ensure that members of a group have had relatively similar life experiences. (Stockard et al., 2009, p. 1456)

> The grouping of cohorts is also of concern, as it determines the number of observations on the group-level. Distinguishing only a few cohorts would not yield enough variation on the cohort-level, increasing the risk of conducting Type-II errors. (Smets & Neundorf, 2014, p. 44)

These diverging justifications indicate a lack of consensus in the applied literature about which cohort selection mechanism produces ideal statistical estimates. Interestingly, selecting a range of cohort years for grouping presents unique problems not faced in other multilevel modeling scenarios. In the case of students nested within schools, for example, researchers cannot *a posteriori* manipulate the student/school nesting structure by altering what schools students belong to. In the HAPC-CCREM, however, the grouping is up for interpretation and to date there is no methodological guidance on that topic.

This conundrum has important implications for statistical estimation of the model, especially in the context of life course sociological theory. There is indeed a tradeoff. Wider birth cohorts, like the ten-year cohort, will have larger sample sizes and greater statistical power to detect cohort-based differences. However, parameter estimates in multilevel modeling are greatly affected by not only the within cluster sample size, but also the number of overall level-two clusters (McNeish & Stapleton, 2016a), an issue that also holds true in cross-classified random effects modeling as well (Luo & Kwok, 2009; Meyers & Beretvas, 2006; Ye & Daniel, 2017). Creating wider birth cohort ranges results in fewer level-two clusters, which may affect model stability (McNeish & Stapleton, 2016a). Although, creating a greater number of level-two clusters by using three- or five-year birth cohorts may result in unequal sample sizes across the cohorts (i.e., unbalanced data), especially for those groups that fall at the tail ends of the sequence of cohorts where cohort sample size is likely much lower than cohorts falling in the middle of the sequence. To address these lingering issues, the second research question for this dissertation is:

> **RQ2**: Even if the cohort selection mechanism used in the estimation of an HAPC-CCREM does not have to match the true cohort selection mechanism underlying the data, does one cohort selection mechanism produce less-biased REML-EB estimates than another?

From a theoretical perspective, the somewhat artificial manipulation of people into these level-two period and cohort groups may be challenging. Societal change, especially on attitudinal measures, takes a long time to enact and become realized. The ten-year cohorts may contain more heterogeneous people and therefore exhibit a wider range of variability in the outcome of interest. However, the three- or five-year cohorts may be more homogenous in the anchoring of historical events that shaped their birth cohort, resulting in less variability in the outcome within these groups. Previous simulation studies have ignored this issue, and to date there has been no examination of the accuracy of estimates from the HAPC-CCREM considering the variability in the period and cohort effects. Taken together, these issues present unique challenges that need further exploration. Statistical simulation studies alongside theoretical development is needed to further explicate how the cohort groupings may impact any detected age, period, or cohort effects, if any detected effects depend on the variability of the period and cohort effects, and how these effects should be interpreted in light of the cohort selection mechanism taken up by the researcher. To answer these important questions, the third research question for this dissertation builds on RQ2 by asking:

> **RQ3**: Does any potential bias exhibited across cohort selection mechanisms depend on the amount of variability in the period or cohort effect?

The last area of investigation for the current study is related to the use of model fit indices in detecting the ideal cohort selection mechanism. Information criteria, such as Akaike's information criteria (AIC) and Bayesian information criterion (BIC), are indices used in the model building process to examine the fit of competing models, where lower values generally represent better fitting models. In practice, researchers typically select as final models those with the lowest value for a particular model fit index. As previously discussed, those using the HAPC-CCREM in practice must make a choice about which cohort selection mechanism to use in the model. Ideally, the cohort selection mechanism used would match the cohort selection mechanism underlying the data, but in reality, we will never know the true cohort selection mechanism underlying the data. Model fit indices, however, may help us determine the cohort selection mechanism that is closest to the truth. In practice, a researcher would specify a baseline HAPC-CCREM only containing predictor variables for the age effect using several competing cohort selection mechanisms. If model fit indices work as intended, the cohort selection mechanism with the lowest fit value would represent the cohort selection mechanism closest to the truth.

The use of model fit indices in this way has not been investigated in the previous simulation studies focused on the HAPC-CCREM. However, for the broader class of cross-classified random effects models in general, Beretvas and Murphy (2013) examined the ability of AIC, BIC, and modifications to these indices to select the best fitting model among a correctly specified model, an under-parameterized model, and an over-specified model. In their simulation study, 48 experimental conditions were examined that varied several experimental factors, including the total sample size, the maximum number of clusters in the cross-classification, the number of non-empty cross-classified cells, and the intraunit correlation coefficient. They found that modifying the AIC to account for the overall sample size and modifying the BIC to account for the number of level-two clusters generally improved the ability to detect the best fitting model. Considering these findings, it is worthwhile to examine the ability of a variety of model fit indices to detect the best fitting model. Accordingly, the final research question for this dissertation is:

> **RQ4**: Given competing cohort selection mechanisms, can model fit indices detect the model that is closest to the true cohort selection mechanism underlying the data?

3 METHODOLOGY

This dissertation aims to investigate two broad areas of methodological issues related to the performance of the hierarchical age-period-cohort cross-classified random effects model (HAPC-CCREM) using a Monte Carlo simulation study. The first area of investigation relates to the accuracy of REML-EB estimates obtained in the HAPC-CCREM. Research questions one through three, outlined in Chapter 2, established four factors that may influence the accuracy of the model estimates: the number of survey years available for analysis, the cohort selection mechanism employed in the model, the variability of the period effect, and the variability of the cohort effect. The second area of investigation in this dissertation, motivated by research question four, is the extent to which model fit indices can be used to help applied researchers in identifying an ideal cohort selection mechanism.

This chapter describes the methodology underlying the Monte Carlo simulation and is structured as follows. First, as a preparatory step for both studies, the results of a real data study analyzing changing attitudes toward abortion are presented. The real data study was a necessary step to generate plausible values for the generating parameters and sampling distributions used in the simulation study (Burton et al., 2006). Following the real data study, separate sections are provided to discuss the study conditions, data generating models, and estimating models relevant to the simulation study design. The chapter concludes with a description of the evaluation criteria associated with the simulation study, which includes model convergence rates, relative absolute bias, coverage rates of the 95% confidence intervals, and analyses of model fit indices.

A Preparatory Step – A Real Data Study

Data Source and Analysis Sample

The real data study utilized the GSS 1972-2016 Cross-Sectional Cumulative Data file (Smith et al., 2017). The GSS was conducted annually from 1972 to 1977, annually or biennially from 1978 to 1993, and biennially from 1994 to 2016. The cumulative data file contained 62,466 respondents across 30 repeated cross-sections. The target population for the 1972 to 2004 survey years was non-institutionalized English-speaking individuals 18 years of age or older living in the United States. From 2006 onward, the target population was expanded to also include Spanish-speaking individuals. GSS provides sampling weights derived from block quota sampling in several of the earlier years and full probability sampling for the majority of the years to extrapolate from the sample to the target population. Following the precedent of previous applications of the HAPC-CCREM to GSS data, the current study was an unweighted analysis.

To arrive at a final sample for analysis, several exclusions were applied to the full sample. The six abortion questions used to construct the dependent variable were not included in the 1986 survey year, resulting in 1,470 individuals (2.35% of the original sample) removed from the analysis. The GSS interview protocol utilizes a split-ballot design to reduce survey administration costs, so that not every participant is asked every question. An additional 45,007 ineligible participants (72.05%) of the original sample were removed because they were not on the ballot to receive the questions used to construct the independent variables. GSS survey protocol allows participants to refuse to answer questions, therefore an additional 2,744 participants (4.39% of the original sample) were removed who refused to answer all questions used to construct the independent variables described below. Finally, 178 participants (0.28% of the original sample) were list-wise deleted because they were missing data on the birth year variable. The final sample size for the real data study was 13,067 individuals (20.92% of the original sample).

Variables of Interest

The dependent variable of interest for the real data study was support for abortion as measured by six abortion ideology questions asked by the GSS. During data collection, participants were asked to report if they thought it should be possible for a pregnant woman to obtain a legal abortion in the following six circumstances: if there is a strong chance of serious defect in the baby; if the woman's own health is seriously endangered by the pregnancy; if she is married and does not want any more children; if the family has a very low income and cannot afford any more children; if she became pregnant as a result of rape; and if she is not married and does not want to marry the man. Responses to the scenarios were coded so that 1 represents yes and 0 represents no. To create the dependent variable, responses to all six scenarios were summed and then multiplied by 100. Therefore, a score of 0 would represent an individual who responded no to all six questions, while a score of 600 would represent an individual who responded yes to all six questions. The dependent variable was scaled in this manner to place the model's parameters in a large enough parameter space to reduce the magnification of otherwise minimal differences in the values of the parameters in the successfully converged models used in the calculation of relative absolute bias (discussed later in this chapter).

Four independent variables were of interest in the real data study: age, sex, the percentage of individuals ever divorced in each cohort, and the percentage of individuals that support divorce laws in each survey year. Respondent age was calculated at the time of the interview by subtracting the respondent's birth year from the year in which they were surveyed. Because the GSS samples adults, the lower limit of the age distribution in the analysis sample was 18. In the data management process, GSS administrators censor the age of respondents who are 90 years of age and older by top-coding their age as 89. Therefore, the last age category represented individuals 89 and older, and technically makes the age variable ordered categorical rather than continuous. However, the age variable was treated as continuous considering that it has 72 categories, each representing an age between 18 and 89 and older. Respondent sex was categorized as male or female, and for the purposes of the current study a dummy indicator was created such that 0 represents males and 1 represents females.

The percentage of individuals ever divorced in each cohort was calculated by aggregating responses of "yes" from the individual-level to the following question: have you ever been divorced or legally separated? Responses were aggregated to the cohort-level using five-year cohorts. The percentage of individuals in each survey year that support divorce laws was calculated by aggregating responses of "easier" to the following question: should divorce in this country be easier or more difficult to obtain than it is now? When such aggregation techniques as these are employed in multilevel models, it is common to calculate contextual effects by leaving the individual-level variable(s) in the model (Raudenbush & Bryk, 2002, pp. 139–141). In the current real data study only the aggregated variables were included to simplify the number of overall parameters examined. Descriptive statistics for all study variables are displayed in Table 1.

Table 1. Descriptive Statistics for Real Data Study Variables

	M	*SD*	*Min*	*Max*
Individual Level				
Abortion	381.85	201.92	0.00	600.00
Age	48.71	16.98	18.00	89.00
Female	0.56	0.50	0.00	1.00
Cohort Level				
Divorce	15.67	6.71	0.00	25.31
Period Level				
Divorce Law	23.30	3.63	17.97	30.25

Note. N = 13,067.

Analysis and Results

Two HAPC-CCREMs were estimated to examine longitudinal changes in support for abortion. First, a baseline model was estimated, which contained linear and quadratic age terms as fixed-effects at level one. Second, a conditional model was estimated that included the age effects from the baseline model and introduced fixed-effects for sex at level one, the cohort-level divorce variable, and the period-level divorce law variable. All models were estimated using restricted maximum likelihood with the lme4 package (Bates et al., 2015) in R version 3.6.0 (R Core Team, 2019). Age was grand-mean centered and scaled by dividing by the standard deviation of age to be within the same parameter space as the other predictor variables. Curvilinear age effects were accounted for by squaring the centered and scaled linear age variable. The period- and cohort-level predictor variables were both grand-mean centered. The female indicator at level one was left uncentered. Across all models, cohorts were grouped using the five-year cohort selection mechanism.

Results of the two HAPC-CCREMs are displayed in Table 2. The baseline APC model that introduced the age effects at level one indicates a negative curvilinear relationship between age and support for abortion. In the conditional model that introduced the individual-, period-, and cohort-level predictors, the negative curvilinear relationship remains. Regarding gender effects, compared to men, women are expected to score 11.071 points lower in support for abortion (p = .002). There was a positive relationship between the prevalence of divorce and support for abortion, such that for every one-point increase in the percentage of divorced individuals in a cohort, support for abortion increased by 3.660 points (p = .002). As the percentage of individuals in a given survey year that believed divorces should be easier to obtain increased, so too did support for abortion (γ_2 = 3.335, p = .002). The additional predictor variables added in the conditional model reduced the amount of explained variability at the individual, period, and cohort levels. However, in both models most of the variability in support for abortion occurred at the individual level.

Table 2. Fixed Effects and Variance Components Estimates for the HAPC-CRREM of Changing Support for Abortion

	Baseline APC	Conditional APC
Fixed Effects		
Intercept (γ_0)	378.37***	373.01***
Age (β_1)	–4.79	–5.93*
Age² (β_2)	–7.75***	–5.74**
Female (β_3)		–11.07**
Divorce (γ_1)		3.66**
Divorce Law (γ_2)		3.34**
Variance Components		
Individual (σ^2)	40120.49	40078.84
Period (τ_v)	303.04	137.94
Cohort (τ_u)	321.92	257.79

Note. N = 13,067. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Simulation Study Design

This simulation study was designed to answer the following research questions proposed in Chapter 2:

RQ1: What is the accuracy of REML-EB estimates at all levels in the HAPC-CCREM given the number of survey years available for analysis?

RQ2: Even if the cohort selection mechanism used in the estimation of an HAPC-CCREM does not have to match the true cohort selection mechanism underlying the data, does one cohort selection mechanism produce less-biased REML-EB estimates than another?

RQ3: Does any potential bias exhibited across cohort selection mechanisms depend on the amount of variability in the period or cohort effect?

RQ4: Given competing cohort selection mechanisms, can model fit indices detect the model that is closest to the true cohort selection mechanism underlying the data?

Simulation Conditions

Answering research question 1 required manipulating the number of available repeated cross-sections for analysis in the HAPC-CCREM. The most recent Monte Carlo simulation study examining the accuracy of REML-EB estimates in the HAPC-CCREM evaluated datasets consisting of 15 cross-sections and a thinned sample of 5 cross-sections (Yang, 2006). This study builds on this previous research by examining the accuracy of REML-EB estimates given 10, 20, and 30 cross-sections of available data. Accordingly, the first study condition consisted of three levels. The 10 cross-sections level was selected as the midpoint between the 5 cross-sections and 15 cross-sections levels used by Yang (2006) to determine if a dataset slightly larger than 5 cross-sections could accurately estimate the model parameters. The levels associated with 20 and 30 cross-sections were selected to reflect real world use of the HAPC-CCREM commonly found in the applied literature (see Appendix A). For example, Meuleman (2019) and Schwadel and Garneau (2014) utilized 22 cross-sections from the GSS to examine changing attitudes towards the redistribution of income and changes in political tolerance, respectively. Additionally, Carlisle and Clark's (2018) analysis of the associations between religion and environmentalism utilized 29 cross-sections from the GSS, while Twenge et al. (2016) utilized 30 cross-sections of GSS data to examine changes in religious participation and beliefs.

Answering research questions 2 and 4 required manipulating the cohort selection mechanism underlying the data. Therefore, the second study condition consisted of three levels, which included the three-year, five-year, and ten-year birth cohort. These condition levels were selected to reflect the cohort selection mechanisms commonly utilized in the applied literature. Answering research question 3 required manipulating the variability in the period and cohort effects. Therefore, the third manipulated condition accounted for the amount of variability in the period effect and has two levels, low period variability and high period variability. Low period variability was operationalized as 2% of the total variability in the study outcome occurring at the period level, while high period variability was operationalized as 10% of the total variability in the study outcome occurring at the period level. These values were designed to reflect the range of period-level variability present in HAPC-CCREM analyses from the applied literature. For example, Anderson et al. (2017) found that 1.67% of the variability in support for the death penalty was at the period level while Schwadel and Ellison (2017) found that 9.82% of the variability in support for the legalization of marijuana was at the period level.

The fourth manipulated condition accounted for the amount of variability in the cohort effect and consisted of two levels, low cohort variability and high cohort variability. Similar to the third study condition, low cohort variability was operationalized as 2% of the total variability in the study outcome occurring at the cohort level, and high cohort variability was operationalized as 10% of the variability in the study outcome occurring at the cohort level. These values were also designed to reflect the range of cohort-level variability found in HAPC-CCREM analyses from the applied literature. For example, in his analysis of changing civic participation, Horowitz (2015) found that 1.91% of the variability in signing a petition was at the cohort level. Additionally, in their analysis of changing political tolerance, Schwadel and Garneau (2014) found that 9.71% of the variability in support for anti-religionists making public speeches was at the cohort level.

In summary, this simulation used a total of four conditions to answer research questions one through four: the number of survey years available for analysis, the cohort selection mechanism underlying the data, the amount of variability in the period effect, and the amount of variability in the cohort effect. To fully explore how these conditions work together to influence the accuracy of estimates in the HAPC-CCREM, a fully-crossed design was utilized. Crossing the levels of all four study conditions resulted in a total of 36 unique study conditions (3 x 3 x 2 x 2), which are displayed in Table 3.

Table 3. Simulation Study Conditions

Manipulated Condition				
Condition	Years	Cohort Selection Mechanism	Period Variability	Cohort Variability
1	10	3-Year	Low	Low
2	20	3-Year	Low	Low
3	30	3-Year	Low	Low
4	10	5-Year	Low	Low
5	20	5-Year	Low	Low
6	30	5-Year	Low	Low
7	10	10-Year	Low	Low
8	20	10-Year	Low	Low
9	30	10-Year	Low	Low
10	10	3-Year	High	Low
11	20	3-Year	High	Low
12	30	3-Year	High	Low
13	10	5-Year	High	Low
14	20	5-Year	High	Low
15	30	5-Year	High	Low
16	10	10-Year	High	Low
17	20	10-Year	High	Low
18	30	10-Year	High	Low
19	10	3-Year	Low	High
20	20	3-Year	Low	High
21	30	3-Year	Low	High
22	10	5-Year	Low	High
23	20	5-Year	Low	High
24	30	5-Year	Low	High
25	10	10-Year	Low	High
26	20	10-Year	Low	High
27	30	10-Year	Low	High
28	10	3-Year	High	High
29	20	3-Year	High	High
30	30	3-Year	High	High
31	10	5-Year	High	High
32	20	5-Year	High	High
33	30	5-Year	High	High
34	10	10-Year	High	High
35	20	10-Year	High	High
36	30	10-Year	High	High

Data Generating Models and Estimating Models

To answer research question 4 regarding the use of model fit indices, baseline models were necessary. Specifically, for each condition, a baseline model was generated such that the expected outcome for individual i in cohort j and period k was calculated as a function of linear and quadratic age effects. The values used in the baseline data generating models for these variables directly correspond to the parameter estimates associated with the fixed effects in the baseline model of the real data study (see Table 2). Note that for all equations in this chapter, superscripts associated with the cohort variance components indicate the chosen value of the cohort selection mechanism, while superscripts associated with the period variance components reflect the presence of single-year periods (that is, that periods were not grouped), borrowing from the notation used by Bell and Jones (2014a).

The sum of the individual, period, and cohort variance components estimates from the baseline model in the real data study (see Table 2) was used to reallocate the amount of variability in the outcome across the individual, period, and cohort levels in the simulation study. These reallocated values, in turn, became the generating parameters used to account for individual-level variability in the outcome, denoted ε_{ijk}, cohort-level variability for the three-year (denoted $u_{0j}^{(3)}$), five-year (denoted $u_{0j}^{(5)}$), and ten-year (denoted $u_{0j}^{(10)}$) cohorts, and period-level variability (denoted $v_{0k}^{(1)}$) in the data generating models. For example, study condition 10 accounts for the scenario in which a 3-year cohort selection mechanism is used in the data generating process, the period effect accounts for 10% of the total variability in the outcome, and the cohort effect accounts for 2% of the total variability in the outcome, leaving 88% of the variability at the individual level. The reallocation of variability in this scenario is therefore calculated as (40,745.70 x 0.88) = 35,856.22 at the individual level, (40,745.70 x 0.10) = 4,074.57 at the period level, and

(40,745.70 x 0.02) = 814.914 at the cohort level. In turn, in the baseline data generating model for condition 10, $\varepsilon_{ijk} \sim N(0, 35856.22)$, $u_{0j}^{(3)} \sim N(0, 814.914)$, and $v_{0k}^{(1)} \sim N(0, 4074.57)$.

To answer research questions one through three regarding the accuracy of estimates in the HAPC-CCREM, conditional models were necessary. Specifically, for each condition, a conditional model was generated such that the expected outcome for individual *i* in cohort *j* and period *k* was calculated as a function of linear and quadratic age effects, level-1 predictor, cohort-level predictor, and period-level predictor. The values used in the conditional data generating models for these variables directly correspond to the parameter estimates associated with the fixed effects in the conditional model of the real data study (see Table 2). Following the same process as the baseline data generating models, the sum of the individual, period, and cohort variance components estimates from the conditional model in the real data study (see Table 2) was used to reallocate the amount of variability in the outcome across the individual, period, and cohort levels in the conditional data generating models.

Across the baseline and conditional data generating models, in study conditions designated for 10 repeated cross-sections, data were generated for 15,000 individuals, while in the 20-year conditions data were generated for 30,000 individuals and in the 30-year conditions data were generated for 45,000 individuals. For each study condition 1,000 datasets were generated, resulting in 36,000 total baseline datasets and 36,000 conditional datasets. Sampling distributions necessary for the data generating models associated with the simulation study are as follows. For both the baseline and conditional data generating models, the individual-level age variable was drawn from a discrete probability distribution designed to match the corresponding distribution in the real data study. Formally, for any level X of age, $0 \leq P(X) \leq 1$ and $\sum P(X) = 1$. The indicator for the level-1 predictor, X, was drawn from the Bernoulli distribution and was designed to match the respondent distribution of the level-1 predictor from the real data study, where:

$$P(X) = \begin{cases} 0.44 & \text{for } X = 0 \\ 0.56 & \text{for } X = 1 \end{cases}. \tag{19}$$

The cohort-level predictor variable, Z, was generated by first drawing from the Bernoulli distribution to specify the individual-level probability for the predictor, matching the individual-level distribution from the real data study, where:

$$P(Z) = \begin{cases} 0.80 & \text{for } Z = 0 \\ 0.20 & \text{for } Z = 1 \end{cases}. \tag{20}$$

Next, cohort-level percentages were calculated by aggregating the data from the individual-level to the cohort-level. The period predictor, W, was generated at the period-level drawing from a four-parameter beta distribution designed to match the period predictor from the real data study, where:

$$W \sim \text{beta}(\alpha = 3, \beta = 8). \tag{21}$$

The beta distribution was bound at the lower end of 18 and the upper end of 30 to match the parameter space of the real data study.

Estimating models for the 36 study conditions were as follows. Across all study conditions, the following set of three baseline estimating models were applied to the baseline datasets, where the outcome was modeled at level-one as a function of linear and quadratic age effects. The first baseline estimating model utilized the three-year cohort selection mechanism, specifically:

$$\begin{gathered} \text{Level 1: } Y_{ijk} = \beta_{0jk} + \beta_1 \text{Age}_{ijk} + \beta_2 \text{Age}_{ijk}^2 + \varepsilon_{ijk}, \varepsilon_{ijk} \sim N(0, \sigma^2) \\ \text{Level 2: } \beta_{0jk} = \gamma_0 + u_{0j}^{(3)} + v_{0k}^{(1)}, u_{0j}^{(3)} \sim N\left(0, \tau_u^{(3)}\right), v_{0k}^{(1)} \sim N(0, \tau_v^{(1)}) \end{gathered} \tag{22}$$

The second baseline estimating model utilized the five-year cohort selection mechanism, specifically:

$$\begin{aligned}\text{Level 1: } Y_{ijk} &= \beta_{0jk} + \beta_1 \text{Age}_{ijk} + \beta_2 \text{Age}_{ijk}^2 + \varepsilon_{ijk}, \varepsilon_{ijk} \sim N(0, \sigma^2) \\ \text{Level 2: } \beta_{0jk} &= \gamma_0 + u_{0j}^{(5)} + v_{0k}^{(1)}, u_{0j}^{(5)} \sim N\left(0, \tau_u^{(5)}\right), v_{0k}^{(1)} \sim N(0, \tau_v^{(1)})\end{aligned} \tag{23}$$

The third baseline estimating model utilized the ten-year cohort selection mechanism, specifically:

$$\begin{aligned}\text{Level 1: } Y_{ijk} &= \beta_{0jk} + \beta_1 \text{Age}_{ijk} + \beta_2 \text{Age}_{ijk}^2 + \varepsilon_{ijk}, \varepsilon_{ijk} \sim N(0, \sigma^2) \\ \text{Level 2: } \beta_{0jk} &= \gamma_0 + u_{0j}^{(10)} + v_{0k}^{(1)}, u_{0j}^{(10)} \sim N\left(0, \tau_u^{(10)}\right), v_{0k}^{(1)} \sim N(0, \tau_v^{(1)})\end{aligned} \tag{24}$$

Note that all baseline estimating models were fit to the data using single-year periods, where $v_{0k}^{(1)} \sim N\left(0, \tau_v^{(1)}\right)$. Across all baseline estimating models, age was grand-mean centered and scaled by dividing by the standard deviation of age. All baseline models were estimated using restricted maximum likelihood with the lme4 package (Bates et al., 2015) in R version 3.6.0 (R Core Team, 2019).

The three, five, and ten-year cohort selection mechanism were selected for use in the baseline estimating models to specifically answer research question 4 concerned with the use of model fit indices to detect an ideal cohort selection mechanism. Applying the three baseline estimating models would mimic the real-world process followed by an applied researcher to consider several competing cohort groupings. Accordingly, across all study conditions, there was one baseline estimating model where the cohort selection mechanism matched the baseline data generating model, and two baseline estimating models where the cohort selection mechanisms did not match.

Across all study conditions, the following set of three conditional estimating models were applied to the conditional datasets, where the outcome was modeled at level-one as a function of linear and quadratic age effects and the individual-level predictor, and at level-two as a function of the cohort-level predictor and period-level predictor. The first conditional estimating model utilized the three-year cohort selection mechanism, specifically:

$$\text{Level 1: } Y_{ijk} = \beta_{0jk} + \beta_1 \text{Age}_{ijk} + \beta_2 \text{Age}^2_{ijk} + \beta_3 X_{ijk} + \varepsilon_{ijk}, \varepsilon_{ijk} \sim N(0, \sigma^2)$$
$$\text{Level 2: } \beta_{0jk} = \gamma_0 + \gamma_1 Z_j + \gamma_2 W_k + u_{0j}^{(3)} + v_{0k}^{(1)},$$
$$u_{0j}^{(3)} \sim N\left(0, \tau_u^{(3)}\right), v_{0k}^{(1)} \sim N(0, \tau_v^{(1)}) \quad (25)$$

The second conditional estimating model utilized the five-year cohort selection mechanism, specifically:

$$\text{Level 1: } Y_{ijk} = \beta_{0jk} + \beta_1 \text{Age}_{ijk} + \beta_2 \text{Age}^2_{ijk} + \beta_3 X_{ijk} + \varepsilon_{ijk}, \varepsilon_{ijk} \sim N(0, \sigma^2)$$
$$\text{Level 2: } \beta_{0jk} = \gamma_0 + \gamma_1 Z_j + \gamma_2 W_k + u_{0j}^{(5)} + v_{0k}^{(1)},$$
$$u_{0j}^{(5)} \sim N\left(0, \tau_u^{(5)}\right), v_{0k}^{(1)} \sim N(0, \tau_v^{(1)}) \quad (26)$$

The third conditional estimating model utilized the ten-year cohort selection mechanism, specifically:

$$\text{Level 1: } Y_{ijk} = \beta_{0jk} + \beta_1 \text{Age}_{ijk} + \beta_2 \text{Age}^2_{ijk} + \beta_3 X_{ijk} + \varepsilon_{ijk}, \varepsilon_{ijk} \sim N(0, \sigma^2)$$
$$\text{Level 2: } \beta_{0jk} = \gamma_0 + \gamma_1 Z_j + \gamma_2 W_k + u_{0j}^{(10)} + v_{0k}^{(1)},$$
$$u_{0j}^{(10)} \sim N\left(0, \tau_u^{(10)}\right), v_{0k}^{(1)} \sim N(0, \tau_v^{(1)}) \quad (27)$$

Note that all conditional estimating models were fit to the data using single-year periods, where $v_{0k}^{(1)} \sim N\left(0, \tau_v^{(1)}\right)$. Across all conditional estimating models, age was grand-mean centered and scaled by dividing by the standard deviation of age. Additionally, the cohort and period predictors were grand mean centered. All conditional models were estimated using restricted maximum likelihood with the lme4 package (Bates et al., 2015) in R version 3.6.0 (R Core Team, 2019).

The three, five, and ten-year cohort selection mechanism were selected for use in the conditional estimating models to specifically answer research question 2, which is partly concerned with accuracy of REML-EB estimates when the cohort selection mechanism used in the estimating model differs from the underlying data. By applying the three estimating models to each study condition, there was one matching and two non-matching cohort selection mechanisms per study condition.

Analyses for Simulation Studies

Research questions one through three are concerned with the accuracy of REML-EB estimates obtained in the HAPC-CCREM. Analyses to assess the accuracy of these estimates across the various study conditions included model convergence rates, relative absolute bias, and coverage rates of the 95% confidence intervals. These analyses are described in more detail below.

Model Convergence Rates

Model convergence rates were evaluated as the percentage of estimated models that failed convergence across the 1,000 iterations associated with each study condition. REML-EB estimation in the HAPC-CCREM is made possible by implementing the expectation-maximization (EM) algorithm. Briefly and conceptually, EM "works" by using an iterative process to derive parameter estimates by choosing the values that maximize the likelihood of the parameters given the observed data (Myung, 2003; Raudenbush & Bryk, 2002, pp. 438–439). When a model does not converge, the EM algorithm fails to find the solution for the maximum likelihood estimate of the parameters in the model, and the model results should not be used.

Relative Absolute Bias

Evaluation of parameter recovery included the intercept, γ_0; the linear and quadratic age effects, β_1 and β_2; the level-one predictor fixed effect, β_3; the level-two cohort predictor fixed effect, γ_1, the level-two period predictor fixed effect, γ_2; and the variance component estimates at the individual, σ^2, period, τ_v, and cohort, τ_u, levels.

Parameter recovery was evaluated using relative absolute bias, defined as:

$$B\left(\bar{\hat{\theta}}_i\right) = \left(\frac{\bar{\hat{\theta}}_i - \theta_i}{|\theta_i|}\right) * 100, \tag{25}$$

where θ_i represents the generated true value of the i^{th} parameter, and $\bar{\hat{\theta}}_i$ represents the average of the estimates $\hat{\theta}_i$ for the i^{th} parameter across *n* successfully converged models per simulation condition. Acceptable boundaries for the value of the relative absolute bias range from −5.00 to 5.00 (Hoogland & Boomsma, 1998). Values outside this range indicate substantial bias, where positive values indicate the overestimation of a parameter and negative values indicate the underestimation of a parameter.

Coverage Rates of the 95% Confidence Intervals

Precision of the fixed effects described above were evaluated using coverage rates of the 95% confidence intervals for each parameter as defined by Morris et al. (2019):

$$\left(\frac{1}{n_{\text{sim}}}\sum_{i=1}^{n_{\text{sim}}}\left(\hat{\theta}_{\text{low},i} \le \theta_i \le \hat{\theta}_{\text{upp},i}\right)\right) * 100\,, \tag{26}$$

where n_{sim} represents the total number of successfully converged models for a given study condition, θ_i represents the generated true value of the i^{th} parameter, $\hat{\theta}_{\text{low},i}$ represents the value of the lower bound of the 95% confidence interval for the estimated i^{th} parameter, and $\hat{\theta}_{\text{upp},i}$ represents the value of the upper bound of the 95% confidence interval for the estimated i^{th} parameter. Lower and upper boundaries for the 95% confidence interval were calculated using the Wald method, specifically:

$$\hat{\theta}_i \pm 1.96 * SE\left(\hat{\theta}_i\right)\,, \tag{27}$$

where $\hat{\theta}_i$ represents the estimated i^{th} parameter and $SE\left(\hat{\theta}_i\right)$ represents the standard error of the estimated i^{th} parameter. Acceptable boundaries for the value of the coverage rates range from 92.5% to 97.5% (Bradley, 1978).

Model Fit Indices to Identify a Cohort Selection Mechanism

The final analysis was specifically related to research question 4 regarding the use of model fit indices to identify the best fitting cohort selection mechanism. Given competing

HAPC-CCREMs estimated using different cohort selection mechanisms, an ideal model fit index would correctly identify the model with the cohort selection mechanism that is either identical or closest to the "true" cohort selection mechanism underlying the data. For example, in study condition 1 (see Table 3), a 3-year cohort selection mechanism was used in the baseline data generating model. As discussed previously, baseline estimating models using the three-, five-, and ten-year cohort selection mechanisms were applied for the study condition. In this scenario, an ideal model fit index would correctly identify the estimating model using the 3-year cohort selection mechanism as the one that best fits the data.

The following model fit indices were specifically evaluated for their ability to identify the best fitting cohort selection mechanism. Akaike's information criterion (Akaike, 1998), or AIC, is calculated as:

$$AIC = -2LL + 2q, \tag{28}$$

where LL represents the value of the log-likelihood of the model upon convergence and q represents the number of fixed effects and variance components in the estimated model. The Bayesian information criterion (Schwarz, 1978), or BIC, is calculated as:

$$BIC = -2LL + \ln(N)\, q, \tag{29}$$

where N represents the number of level-one units used in the estimated model. The AIC and BIC model fit indices were selected for evaluation because they are readily available in statistical software and were commonly employed in the applied literature using the HAPC-CCREM reviewed in Chapter 2.

Previous research on multilevel models in general (Gurka, 2006) and cross-classified random effects models in particular (Beretvas & Murphy, 2013) have shown that modifications to AIC and BIC improve the ability of the indices to select the best fitting model. Therefore, the following modified indices were also evaluated. AIC was modified to correct for sample size using the finite sample-corrected AIC (AICC, Hurvich & Tsai, 1989), and is calculated as

$$AICC = \frac{-2LL + (2qN)}{(N - q - 1)}. \tag{30}$$

Given that BIC already accounts for sample size, it will be modified to account for the number of level-two units in the analysis:

$$BICC = -2LL + \ln(M)\, q, \tag{31}$$

where *M* represents the number of cross-classified period by cohort cells present in the data used by the model. Performance of the model fit indices described above were evaluated as the percentage of correct model identifications made for the *n* successfully converged models associated with each study condition.

Scientific Reproducibility

Appendix B contains the R program written to conduct the real data study and Monte Carlo simulation described in this chapter. For the Monte Carlo simulation study, seeds are set where necessary to ensure the random distributions used in the data generating models are fully reproducible (Morris et al., 2019).

4 RESULTS

This chapter describes the results of the Monte Carlo simulation study designed to assess the accuracy of estimates in hierarchical age-period-cohort models of repeated cross-sectional data as well as the extent to which model fit indices can be used to help applied researchers in identifying an ideal cohort selection mechanism. Recall from Chapter 3 that four factors are under investigation: the number of repeated cross-sections available for analysis (10, 20, and 30 years), the cohort selection mechanism employed (3-year, 5-year, and 10-year), the amount of variability for the period clusters (low and high), and the amount of variability for the cohort clusters (low and high). The fully crossed study design results in a total of 36 unique study conditions, and for each simulation outcome results are presented tabularly along these lines. As discussed in Chapter 3, for both the baseline and conditional models, 1,000 datasets were generated, resulting in 72,000 total datasets.

The structure of this chapter is organized as follows. First, the percentage of estimating models that failed maximum likelihood convergence is discussed. Second, the implementation of model fit indices to identify the correct cohort selection mechanism used to generate the data is discussed. Third, relative absolute bias and coverage rates of the 95% confidence intervals are described for each of the following fixed effects parameters: the model intercept, linear and curvilinear age terms, the level-1 predictor variable, the level-2 cohort predictor variable, and the level-2 period predictor variable. Finally, relative absolute bias for the level-1 individual variance component, level-2 period variance component, and level-2 cohort variance component is discussed.

Model Convergence Rates

The percentage of models that failed convergence are displayed in Table 4 by study condition and estimating model. There were no study conditions in which all estimating models

Table 4. Percentage of Models across the 1,000 Replications that Failed Convergence by Study Condition and Estimating Model.

Manipulated Condition					CSM in Estimating Model		
Condition	Years	CSM DGM	Period Variability	Cohort Variability	3-Year	5-Year	10-Year
1	10	3-Year	Low	Low	7.1	4.3	10.4
2	20	3-Year	Low	Low	8.0	5.7	10.3
3	30	3-Year	Low	Low	6.6	4.5	8.5
4	10	5-Year	Low	Low	7.1	5.9	9.3
5	20	5-Year	Low	Low	7.0	6.7	11.9
6	30	5-Year	Low	Low	6.5	4.8	10.7
7	10	10-Year	Low	Low	7.4	6.3	6.5
8	20	10-Year	Low	Low	5.1	6.4	5.8
9	30	10-Year	Low	Low	7.0	4.8	7.6
10	10	3-Year	High	Low	4.3	4.9	8.1
11	20	3-Year	High	Low	6.5	5.2	10.5
12	30	3-Year	High	Low	4.7	3.4	6.8
13	10	5-Year	High	Low	5.7	6.1	10.9
14	20	5-Year	High	Low	4.6	5.6	12.2
15	30	5-Year	High	Low	4.9	4.1	10.2
16	10	10-Year	High	Low	4.8	3.6	5.3
17	20	10-Year	High	Low	5.1	5.9	6.3
18	30	10-Year	High	Low	4.8	3.8	8.1
19	10	3-Year	Low	High	9.8	7.5	9.1
20	20	3-Year	Low	High	7.6	7.3	11.0
21	30	3-Year	Low	High	4.3	4.3	10.0
22	10	5-Year	Low	High	8.8	7.0	9.3
23	20	5-Year	Low	High	7.0	5.3	8.7
24	30	5-Year	Low	High	5.4	5.0	8.3
25	10	10-Year	Low	High	9.6	7.9	5.6
26	20	10-Year	Low	High	6.8	5.8	6.6
27	30	10-Year	Low	High	5.0	3.7	5.7
28	10	3-Year	High	High	11.4	6.4	9.3
29	20	3-Year	High	High	9.2	6.9	11.3
30	30	3-Year	High	High	6.4	7.9	10.9
31	10	5-Year	High	High	9.2	4.3	10.7
32	20	5-Year	High	High	6.2	5.5	12.5
33	30	5-Year	High	High	7.0	4.4	12.0
34	10	10-Year	High	High	7.9	5.6	4.3
35	20	10-Year	High	High	6.1	6.2	7.5
36	30	10-Year	High	High	5.9	4.2	7.5

Note. Shaded table cells indicate conditions where there is a match between the cohort selection mechanism in the data generating and estimating models. CSM DGM = cohort selection mechanism used in the data generating model.

successfully converged. The estimating model with the smallest percentage of models (3.4%) that failed convergence was the 5-year cohort selection mechanism used in condition 12, where the data was generated using a 3-year cohort selection mechanism with high period variability, low cohort variability, and thirty years of data. The estimating model with the highest percentage of models (12.5%) that failed convergence was the 10-year cohort selection mechanism used in condition 32, where the data was generated using a 5-year cohort selection mechanism with high period variability, high cohort variability, and twenty years of data.

In 29 of the 36 conditions (81%), the 10-year estimating model had a higher percentage of models that failed convergence compared to the 3-year and 5-year estimating models. The seven conditions that did not follow this pattern tended to be scenarios in which a 10-year cohort selection mechanism was used to generate the data, and the 3-year estimating model had a higher percentage of models that failed convergence compared to the 5-year and 10-year estimating models (for example, see conditions 7, 25, 26, and 34). One notable exception was condition 28, where 10 years of data were generated using a 3-year cohort selection mechanism with high period and cohort variability. In this condition, the 3-year estimating model had the greatest percentage of models that failed convergence (11.40%).

In 27 of the 36 conditions (75%), the 5-year estimating model had the lowest percentage of models that failed converge across the estimating models, which may help to explain why five-year cohorts were so commonly utilized in the applied literature (see Appendix A). In four of the nine conditions that did not follow this pattern, the estimating model with the lowest percentage of models that failed convergence were those where the cohort selection mechanism used in the data generating model matched the cohort selection mechanism used in the estimating model (see conditions 10, 25, 30, and 34). In the remaining five conditions, the 3-year estimating model had the lowest percentage of models that failed convergence (see conditions 8, 13, 14, 17, and 35), although the accompanying data generating model used either the 5-year or 10-year cohort selection mechanism.

There was no readily apparent relationship between the number of repeated cross-sections used in the data generating model and the percentage of models that failed convergence. In fact, in many of the study conditions, the percentage of models that failed convergence was greater in datasets generated with 20 repeated cross-sections compared to their corresponding study conditions generated with 10 or 30 repeated cross-sections where the remaining condition levels were identical (for example, see conditions 1, 2, and 3). Generally, there appeared to be a greater difference in the percentage of models that failed convergence between the cohort selection mechanism utilized in the estimating model compared to how many repeated cross-sections were used to generate the difference. For example, in study conditions 4, 5, and 6, the difference between the percentage of models that failed to converge was upwards of five percentage points when comparing the cohort selection mechanisms in the estimating model within a given condition. When comparing the number of survey years in the data generating model (i.e., moving across each condition but examining the percentage using the same cohort selection mechanism used in the estimating model), these differences decrease to one to two percentage points.

Model Fit Indices to Identify a Cohort Selection Mechanism

Correct model identification rates by model fit index are presented in Table 5 by study condition. Cell values under each model fit index column header represent the percentage of successfully converged models where the corresponding fit index correctly identified the estimating model with the cohort selection mechanism that was used to generate the data. All of the model fit indexes performed very well. In 29 of the 36 study conditions (81%), all of the fit indexes correctly identified the cohort selection mechanism used in the data generating model for

Table 5. Correct Model Identification Rates by Model Fit Index and Study Condition

Manipulated Condition					Model Fit Index			
Condition	Years	CSM DGM	Period Variability	Cohort Variability	AIC	BIC	AICC	BICC
1	10	3-Year	Low	Low	100.00	100.00	100.00	100.00
2	20	3-Year	Low	Low	100.00	100.00	100.00	100.00
3	30	3-Year	Low	Low	100.00	100.00	100.00	100.00
4	10	5-Year	Low	Low	99.37	99.37	99.37	99.37
5	20	5-Year	Low	Low	99.74	99.74	99.74	99.74
6	30	5-Year	Low	Low	100.00	100.00	100.00	100.00
7	10	10-Year	Low	Low	98.77	98.77	98.77	99.75
8	20	10-Year	Low	Low	99.88	99.88	99.88	100.00
9	30	10-Year	Low	Low	100.00	100.00	100.00	100.00
10	10	3-Year	High	Low	100.00	100.00	100.00	100.00
11	20	3-Year	High	Low	100.00	100.00	100.00	100.00
12	30	3-Year	High	Low	100.00	100.00	100.00	100.00
13	10	5-Year	High	Low	99.62	99.62	99.62	99.49
14	20	5-Year	High	Low	100.00	100.00	100.00	100.00
15	30	5-Year	High	Low	100.00	100.00	100.00	100.00
16	10	10-Year	High	Low	99.54	99.54	99.54	100.00
17	20	10-Year	High	Low	99.76	99.76	99.76	99.88
18	30	10-Year	High	Low	100.00	100.00	100.00	100.00
19	10	3-Year	Low	High	100.00	100.00	100.00	100.00
20	20	3-Year	Low	High	100.00	100.00	100.00	100.00
21	30	3-Year	Low	High	100.00	100.00	100.00	100.00
22	10	5-Year	Low	High	100.00	100.00	100.00	100.00
23	20	5-Year	Low	High	100.00	100.00	100.00	100.00
24	30	5-Year	Low	High	100.00	100.00	100.00	100.00
25	10	10-Year	Low	High	100.00	100.00	100.00	100.00
26	20	10-Year	Low	High	100.00	100.00	100.00	100.00
27	30	10-Year	Low	High	100.00	100.00	100.00	100.00
28	10	3-Year	High	High	100.00	100.00	100.00	100.00
29	20	3-Year	High	High	100.00	100.00	100.00	100.00
30	30	3-Year	High	High	100.00	100.00	100.00	100.00
31	10	5-Year	High	High	100.00	100.00	100.00	100.00
32	20	5-Year	High	High	100.00	100.00	100.00	100.00
33	30	5-Year	High	High	100.00	100.00	100.00	100.00
34	10	10-Year	High	High	100.00	100.00	100.00	100.00
35	20	10-Year	High	High	100.00	100.00	100.00	100.00
36	30	10-Year	High	High	100.00	100.00	100.00	100.00

Note. CSM DGM = cohort selection mechanism used in the data generating model.

100% of the successfully converged models. In the remaining seven conditions, the performance of the indexes only slightly decreased, usually between one to two percentage points. In three study conditions (conditions 8, 16, and 17), the BICC index marginally outperformed the other indexes. The common factor across these three conditions was that the 10-year cohort selection mechanism was used to generate the data. In two study conditions, the AIC, BIC, and AICC indexes marginally outperformed the BICC index. The commonalities in the data generating models for these two study conditions were that 10 repeated cross-sections were used to generate the data with low variability placed on the cohort clusters.

Relative Absolute Bias and Coverage Rates of Fixed Effects Estimates

Intercept Parameter, γ_0

Relative absolute bias and coverage rates of the intercept parameter, γ_0, are displayed in Table 6. Across all study conditions the intercept was recovered within the acceptable boundaries for relative absolute bias. In fact, in only one condition was relative absolute bias greater than half a percentage point. In condition 34, which was generated using 10 repeated cross-sections, a 10-year cohort selection mechanism, and high variability placed on the period and cohort clusters, the intercept was overestimated between 0.50% to 0.66% across the three estimating models but this was still within the acceptable threshold (Hoogland & Boomsma, 1998).

Coverage of the intercept appeared to be mainly driven by the cohort selection mechanism in the estimating model matching the data generating model. Generally, when the cohort selection mechanism in the estimating model matched the data generating model, the intercept was appropriately covered, ranging from 92.51% to 95.53%. The exception to this pattern, which occurred for several conditions, was when the cohort selection mechanism in the estimating model matched the data generating model and the data was generated using 10 repeated cross-sections (see conditions 7, 10, 13, 16, and 34). In all of these scenarios, the intercept was slightly

Table 6. Relative Absolute Bias (RAB) and Coverage Rates of Intercept Parameter, γ_0

Manipulated Condition					RAB by CSM in Estimating Model			Coverage Rates by CSM in Estimating Model		
Con.	Years	CSM in DGM	Period Var.	Cohort Var.	3-Year	5-Year	10-Year	3-Year	5-Year	10-Year
1	10	3-Year	Low	Low	–0.08	–0.03	–0.03	93.72	**91.84**	**90.25**
2	20	3-Year	Low	Low	–0.03	–0.10	–0.11	94.31	93.12	93.68
3	30	3-Year	Low	Low	–0.06	–0.21	–0.08	94.93	93.86	94.07
4	10	5-Year	Low	Low	0.01	0.13	–0.03	**88.62**	94.07	**90.22**
5	20	5-Year	Low	Low	0.06	0.02	–0.01	**85.31**	93.29	93.22
6	30	5-Year	Low	Low	–0.12	–0.06	–0.05	**87.13**	94.63	94.18
7	10	10-Year	Low	Low	0.17	0.03	0.16	**80.02**	**86.38**	**92.13**
8	20	10-Year	Low	Low	0.07	–0.01	0.06	**82.77**	**88.53**	94.28
9	30	10-Year	Low	Low	0.04	–0.25	–0.08	**79.45**	**84.59**	**92.02**
10	10	3-Year	High	Low	–0.17	–0.24	–0.31	**91.99**	**91.02**	**92.07**
11	20	3-Year	High	Low	0.13	0.07	0.08	94.29	93.45	93.90
12	30	3-Year	High	Low	–0.06	–0.11	–0.03	93.80	93.17	94.25
13	10	5-Year	High	Low	0.07	–0.11	0.20	**89.60**	**91.75**	**92.37**
14	20	5-Year	High	Low	–0.08	–0.04	–0.03	**90.84**	93.87	94.43
15	30	5-Year	High	Low	–0.02	–0.05	–0.01	**92.48**	94.45	96.63
16	10	10-Year	High	Low	–0.16	–0.17	–0.16	**86.55**	**87.95**	**92.00**
17	20	10-Year	High	Low	–0.17	–0.40	–0.25	**89.50**	**91.12**	93.91
18	30	10-Year	High	Low	0.19	0.10	0.23	**90.33**	**91.96**	93.17
19	10	3-Year	Low	High	0.21	0.12	0.01	95.21	92.54	**86.80**
20	20	3-Year	Low	High	0.05	–0.02	–0.10	94.73	94.04	96.16
21	30	3-Year	Low	High	–0.20	–0.32	–0.29	93.20	**92.29**	96.61
22	10	5-Year	Low	High	–0.14	–0.12	0.06	**85.78**	94.29	**87.67**
23	20	5-Year	Low	High	0.34	0.35	0.40	**84.20**	93.16	95.15
24	30	5-Year	Low	High	0.05	0.08	0.11	**84.34**	95.07	**98.11**
25	10	10-Year	Low	High	0.30	0.14	0.10	**74.73**	**85.19**	93.18
26	20	10-Year	Low	High	0.08	0.03	0.07	**72.25**	**82.96**	92.59
27	30	10-Year	Low	High	0.08	0.02	0.10	**71.43**	**83.39**	**92.46**
28	10	3-Year	High	High	–0.01	0.10	0.22	92.51	**91.45**	93.05
29	20	3-Year	High	High	0.13	–0.11	–0.05	94.93	94.77	95.22
30	30	3-Year	High	High	–0.17	–0.18	0.00	95.53	96.32	96.89
31	10	5-Year	High	High	0.01	0.26	–0.05	**89.97**	93.70	**90.47**
32	20	5-Year	High	High	0.08	0.05	0.08	**87.78**	94.58	96.85
33	30	5-Year	High	High	0.07	0.04	0.20	**88.46**	93.79	**97.66**
34	10	10-Year	High	High	0.66	0.52	0.50	**83.44**	**87.22**	**92.39**
35	20	10-Year	High	High	–0.15	–0.31	–0.18	**83.14**	**87.79**	94.19
36	30	10-Year	High	High	–0.17	–0.21	–0.17	**79.68**	**85.73**	93.49

Note. Items in bold indicate results outside of the acceptable boundaries for relative absolute bias or coverage. Shaded table cells indicate conditions where there is a match between the cohort selection mechanism in the data generating and estimating models. Con. = study condition; CSM = cohort selection mechanism; DGM = data generating model; Var. = variability.

under-covered (range of coverage rates = 91.75% to 92.39%). There were two additional conditions where the cohort selection mechanism in the estimating model matched the data generating model, but the intercept was slightly under-covered. These were conditions 9 (coverage = 92.02%) and 27 (coverage = 92.46%), both of which were generated using 30 repeated cross-sections and the 10-year cohort selection mechanism in the data generating model.

An additional finding regarding coverage of the intercept was as follows. In study conditions where the data were generated using the 3-year cohort selection mechanism, the intercept was frequently covered appropriately by the estimating models using the 5-year or 10-year cohort selection mechanism (see conditions 28, 29, and 30). However, in the conditions where the data were generated using the 5-year cohort selection mechanism, the estimating model using the 3-year cohort selection mechanism always under-covered the intercept by 84.20% to 92.48% (for example, see conditions 31, 32, and 33). Interestingly, when the data were generated using 10-year cohorts, the estimating models using 3-year and 5-year cohorts were always under-covered (for example, see conditions 16, 17, and 18), and the degree of under-coverage was more substantial (range = 71.43% to 90.33% for 3-year estimating models and 82.96% to 91.96% for 5-year estimating models). Also of note is that only two scenarios existed where the intercept was over-covered. These occurred in the 10-year estimating model in conditions 24 (coverage = 98.11%) and 33 (coverage = 97.66%), each of which had data generated using 30 repeated cross-sections with a 5-year cohort selection mechanism. Across all study conditions, there was no readily apparent relationship between the amount of variability at the period and cohort levels and coverage of the intercept.

Level-1 Linear Age Parameter, $\boldsymbol{\beta_1}$

Relative absolute bias and coverage rates of the level-1 linear age parameter, β_1, are displayed in Table 7. Regarding relative absolute bias, there was one condition for which the

Table 7. Relative Absolute Bias (RAB) and Coverage Rates of Level-1 Linear Age Parameter, β_1

Manipulated Condition					RAB by CSM in Estimating Model			Coverage Rates by CSM in Estimating Model		
Con.	Years	CSM in DGM	Period Var.	Cohort Var.	3-Year	5-Year	10-Year	3-Year	5-Year	10-Year
1	10	3-Year	Low	Low	–2.49	–4.21	**–8.17**	95.74	**88.45**	**62.59**
2	20	3-Year	Low	Low	1.76	0.26	0.62	94.63	**85.70**	**52.99**
3	30	3-Year	Low	Low	–0.60	–0.93	**–7.69**	96.52	**88.55**	**59.78**
4	10	5-Year	Low	Low	–4.86	–4.07	1.01	**82.61**	93.23	**44.73**
5	20	5-Year	Low	Low	–2.94	–3.03	**9.00**	**83.91**	94.99	**38.56**
6	30	5-Year	Low	Low	–1.66	–3.16	–0.30	**85.15**	95.16	**40.11**
7	10	10-Year	Low	Low	**–7.68**	**–8.44**	–3.39	**65.59**	**80.15**	**92.03**
8	20	10-Year	Low	Low	0.46	1.28	–1.15	**67.65**	**80.00**	**91.21**
9	30	10-Year	Low	Low	1.14	1.07	0.44	**68.81**	**82.60**	92.96
10	10	3-Year	High	Low	–2.09	**–6.84**	**–8.16**	94.31	**86.20**	**58.92**
11	20	3-Year	High	Low	–0.07	0.41	–0.22	94.81	**85.41**	**52.66**
12	30	3-Year	High	Low	1.75	1.58	**–5.13**	94.23	**84.44**	**52.71**
13	10	5-Year	High	Low	–2.81	–3.57	–2.70	**81.10**	93.03	**43.88**
14	20	5-Year	High	Low	–1.49	–2.74	**–9.25**	**81.48**	**92.18**	**33.85**
15	30	5-Year	High	Low	**–5.72**	**–5.50**	**–25.17**	**82.78**	94.45	**34.79**
16	10	10-Year	High	Low	2.35	2.45	1.08	**63.88**	**79.02**	**91.47**
17	20	10-Year	High	Low	–2.45	–2.49	–2.96	**68.80**	**83.19**	92.83
18	30	10-Year	High	Low	–1.23	–1.39	1.43	**67.09**	**81.73**	93.17
19	10	3-Year	Low	High	0.94	–4.39	**–34.86**	93.91	**56.93**	**26.61**
20	20	3-Year	Low	High	**–6.49**	**5.31**	**51.12**	94.19	**53.25**	**21.30**
21	30	3-Year	Low	High	–1.48	**–12.44**	**36.85**	94.25	**54.70**	**22.21**
22	10	5-Year	Low	High	**–10.40**	–4.92	–0.85	**76.08**	92.81	**16.74**
23	20	5-Year	Low	High	1.74	–1.74	**–71.60**	**74.13**	94.84	**13.55**
24	30	5-Year	Low	High	**–5.29**	0.10	**–73.84**	**76.83**	93.50	**13.43**
25	10	10-Year	Low	High	**11.76**	**9.69**	4.35	**64.93**	**83.28**	**92.34**
26	20	10-Year	Low	High	**7.05**	–0.48	–4.15	**63.03**	**83.71**	92.70
27	30	10-Year	Low	High	4.62	**5.32**	1.73	**69.66**	**87.88**	93.31
28	10	3-Year	High	High	0.07	**–40.54**	**–18.61**	94.22	**55.97**	**24.70**
29	20	3-Year	High	High	–1.72	**31.81**	**40.92**	93.96	**42.70**	**21.98**
30	30	3-Year	High	High	–2.15	**26.55**	**15.58**	95.10	**44.85**	**20.56**
31	10	5-Year	High	High	**8.75**	**14.24**	–3.20	**74.39**	93.60	**18.29**
32	20	5-Year	High	High	–0.12	0.81	**–12.46**	**70.20**	94.69	**14.27**
33	30	5-Year	High	High	–4.09	–3.40	**–22.33**	**72.06**	93.16	**11.82**
34	10	10-Year	High	High	**–14.50**	**–8.90**	–2.27	**63.35**	**84.03**	93.01
35	20	10-Year	High	High	–3.72	1.16	0.18	**59.54**	**80.21**	94.30
36	30	10-Year	High	High	**6.96**	**6.19**	1.53	**59.37**	**84.68**	95.55

Note. Items in bold indicate results outside of the acceptable boundaries for relative absolute bias or coverage. Shaded table cells indicate conditions where there is a match between the cohort selection mechanism in the data generating and estimating models. Con. = study condition; CSM = cohort selection mechanism; DGM = data generating model; Var. = variability.

linear age parameter was estimated with bias across all three estimating models. In study condition 15, where data were generated using 30 repeated cross-sections using a 5-year cohort selection mechanism with high period variability and low cohort variability, all three estimating models underestimated the linear age parameter (6% underestimation for the 3-year and 5-year cohort selection mechanisms, and 25% underestimation for the 10-year). Otherwise, there appeared to be two determinants of relative absolute bias for the linear age parameter. The first was the extent to which the cohort selection mechanism in the data generating model matched the estimating model. Generally, when the two matched, the linear age parameter was recovered without bias. The second determinant occurred in study conditions where the cohort selection mechanisms between the data generating and data estimating models did not match and cohorts were generated with high variability (conditions 19 through 36). In almost all of these study conditions, the 10-year estimating model recovered the linear age parameter with bias, underestimating by as much as 74% (condition 24) and overestimating by as much as 51% (condition 20). This finding also extended to the 5-year estimating models when high variability was placed on the periods in addition to the cohorts (conditions 28 through 36), where the linear age parameter was underestimated by as much as 41% and overestimated by as much as 32%.

The impact of these factors on relative absolute bias for the linear age parameter is displayed in Figure 1. Figure 1 shows the dispersion of the estimated linear age parameter by study condition and estimating model for conditions 3, 21, and 30. Each blue point represents the estimate of the age parameter for each successfully converged model among each cohort selection mechanism, and the red vertical line represents the mean of the estimates among each cohort selection mechanism. Each of the three conditions in the figure were generated using 30-repeated cross sections of data with a 3-year cohort selection mechanism. In condition 3, where low

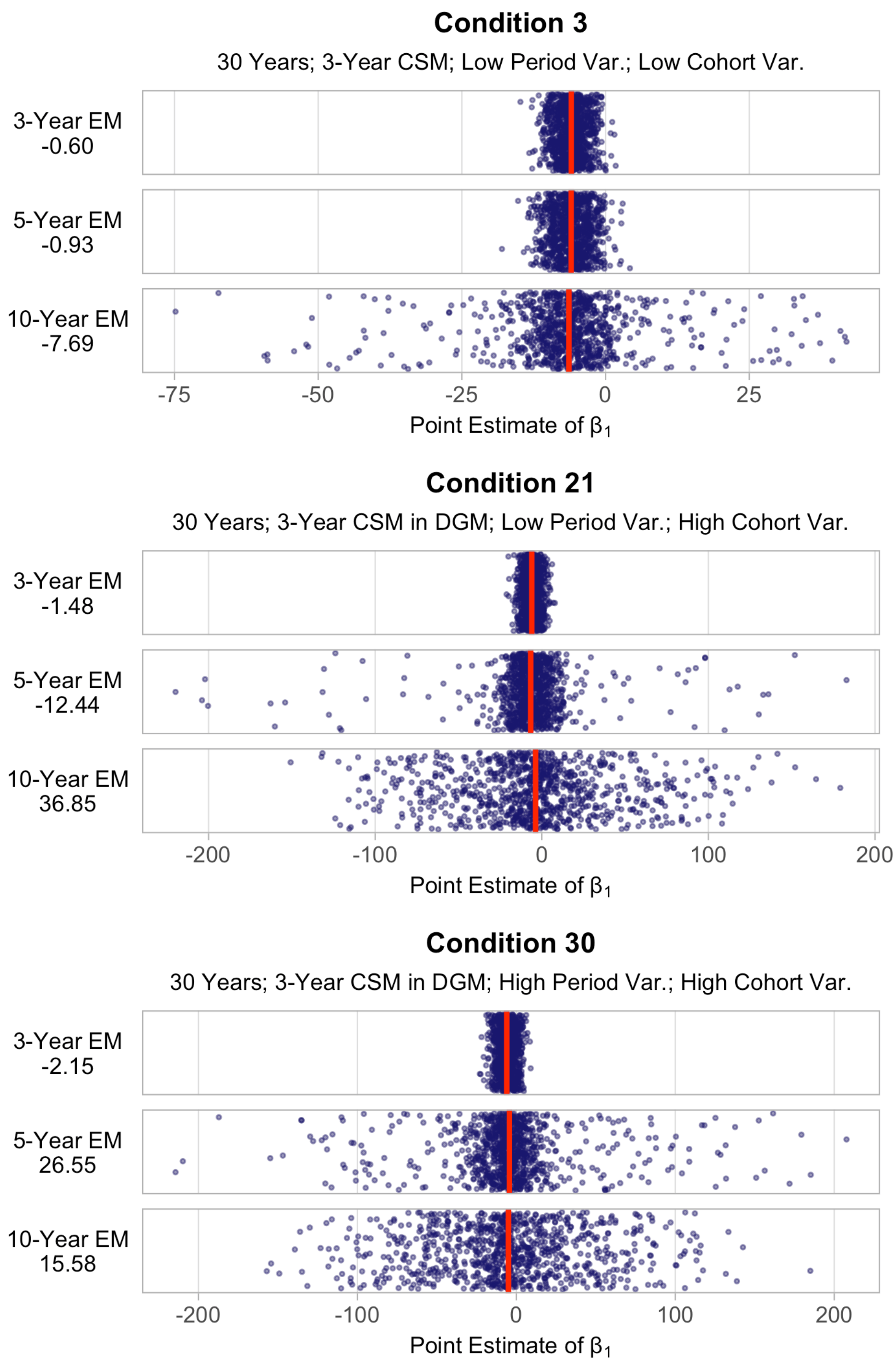

Figure 1. Dispersion Plot of Level-1 Linear Age Parameter, β_1, for Three Study Conditions

Note. Blue points represent the estimate of the age parameter for each successfully converged model. Red vertical lines represent the mean of the estimates among each cohort selection mechanism. Values under the estimating model label indicate the value of the relative absolute bias. CSM = cohort selection mechanism; DGM = data generating model; Var. = variability; EM = estimating model.

variability was placed on the period and cohorts, only the 10-year estimating model exhibited parameter bias, underestimating the linear age effect by 8%. When high variability is placed on the cohorts, as in condition 21, the magnitude of relative absolute bias increases for the non-matching cohort selection mechanisms. This pattern holds true when high variability is placed on both the period and cohorts, as exhibited in the plot for condition 30, which visualizes the wider dispersion of the age parameter for the 5-year and 10-year estimating models, which overestimated the age parameter by 27% and 16%, respectively.

Coverage rates of the 95% confidence interval for the level-1 linear age parameter also appeared to be driven by mismatching cohort selection mechanisms between the data generating and data estimating models and the amount of variability placed on the period and cohort clusters. Generally, when the cohort selection mechanisms matched, the age effect was appropriately covered. One notable exception was the 10-year cohort selection mechanism. There were four conditions where the 10-year cohort selection mechanism consistently under-covered the age parameter even though the data were generated using 10-year cohorts (see conditions 7, 8, 16, and 25). In these four conditions, the linear age parameter was slightly under-covered, ranging from 91.21% to 92.34%. The issue appears to be influenced by the number of repeated cross-sections used to generate the data; in three of the four conditions, the data were generated using the smallest amount of (i.e., 10) repeated cross-sections. Across all of the study conditions, when the cohort selection mechanisms used in the data generating models and estimating models did not match, the age parameter was always under-covered. For 5-year estimating models that did not match the generating models, coverage rates ranged from 42.70% to 88.55%. The magnitude of the under-coverage was particularly high for the 10-year estimating models (coverage rates = 11.82% to 62.59%), especially when high variability was placed on the cohorts (conditions 19 through 36).

The impact of these factors on coverage rates for the linear age parameter is displayed as a zipper plot in Figure 2. In the figure, grey vertical lines represent the true parameter value for the linear age parameter used in the data generating model. Each red horizontal line represents one 95% confidence interval for a successfully converged model that did not contain the true linear age parameter, while each blue horizontal line indicates a confidence interval that did contain the true parameter value. Each of the three study conditions in the figure were generated using 30 repeated cross-sections of data with a 5-year cohort selection mechanism. Across all conditions in the figure, when 5-year cohorts were used in the estimating model the linear age parameter was sufficiently covered. However, in condition 6, where the periods and cohorts were generated with low variability, both the 3-year and 10-year estimating models under-covered the age parameter (coverage = 85.15% and 40.11%, respectively). Coverage rates for the 3-year and 5-year estimating models decreased further when cohorts were generated with high variability (condition 24, coverage = 76.83% and 13.43%). Finally, when both period and cohorts were generated with high variability, coverage of the 10-year estimating model drastically decreased to 11.82.

Level-1 Curvilinear Age Parameter, β_2

Relative absolute bias and coverage rates of the level-1 curvilinear age parameter, β_2, are displayed in Table 8. There were four study conditions where the level-1 curvilinear age parameter was estimated with moderate bias (conditions 19, 22, 25, and 28). There were two common factors across these four conditions: the data were generated using 10 repeated cross-sections and the cohorts were generated with high variability. In three of the four conditions, a 10-year cohort selection mechanism was applied in the estimating model that did not match the cohort selection mechanism used in the data generating model, resulting in underestimation as low as 6% and

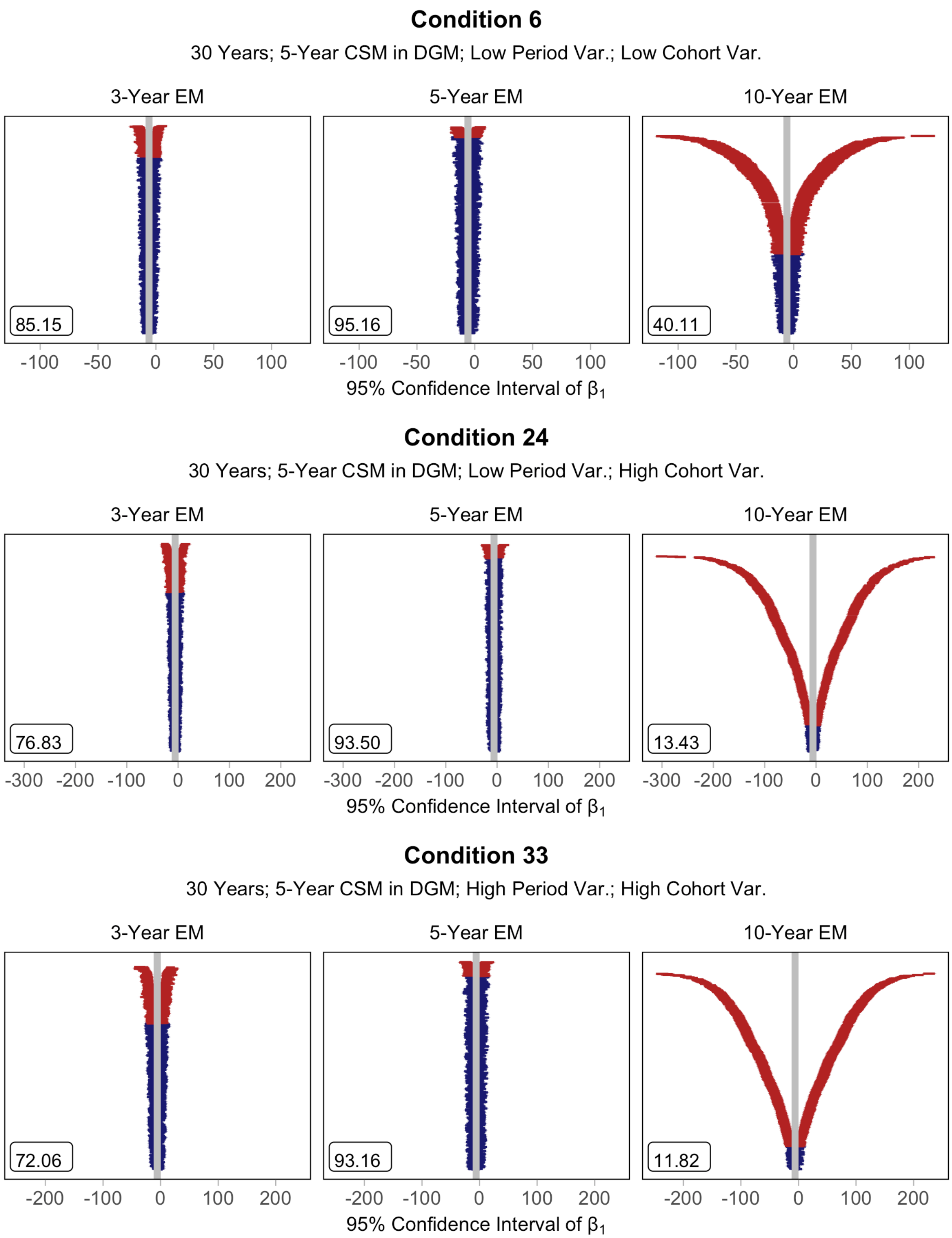

Figure 2. Zipper Plot of Level-1 Linear Age Parameter, β_1, for Three Study Conditions

Note. CSM = cohort selection mechanism; DGM = data generating model; Var. = variability; EM = estimating model.

Table 8. Relative Absolute Bias (RAB) and Coverage Rates of Level-1 Curvilinear Age Parameter, β_2

Manipulated Condition					RAB by CSM in Estimating Model			Coverage Rates by CSM in Estimating Model		
Con.	Years	CSM in DGM	Period Var.	Cohort Var.	3-Year	5-Year	10-Year	3-Year	5-Year	10-Year
1	10	3-Year	Low	Low	2.82	3.06	2.47	94.25	**88.24**	**68.27**
2	20	3-Year	Low	Low	0.09	–0.77	1.27	95.60	93.66	**78.27**
3	30	3-Year	Low	Low	–0.15	–0.37	0.06	95.25	94.17	**90.11**
4	10	5-Year	Low	Low	–0.61	–2.99	0.90	**85.46**	94.92	**59.23**
5	20	5-Year	Low	Low	0.32	0.12	0.93	93.30	95.53	**73.33**
6	30	5-Year	Low	Low	0.19	0.73	–0.19	94.56	95.05	**88.35**
7	10	10-Year	Low	Low	–3.06	–1.01	–3.35	**73.74**	**87.22**	93.91
8	20	10-Year	Low	Low	–0.32	–0.22	–0.34	**89.01**	92.84	94.60
9	30	10-Year	Low	Low	–0.86	–0.54	–0.81	**92.41**	94.13	94.96
10	10	3-Year	High	Low	1.33	3.73	4.22	93.47	**86.52**	**64.76**
11	20	3-Year	High	Low	–0.66	–1.20	–1.25	95.34	**92.07**	**77.83**
12	30	3-Year	High	Low	–0.43	–0.35	–0.02	94.55	94.53	**91.49**
13	10	5-Year	High	Low	–3.53	–2.31	–2.72	**87.05**	93.25	**55.78**
14	20	5-Year	High	Low	–0.66	–0.29	0.09	93.03	95.77	**71.27**
15	30	5-Year	High	Low	0.46	0.29	0.18	93.32	94.55	**84.96**
16	10	10-Year	High	Low	2.33	3.43	3.34	**75.32**	**89.20**	93.60
17	20	10-Year	High	Low	–0.35	0.35	1.10	**88.55**	93.23	94.13
18	30	10-Year	High	Low	0.11	0.58	0.74	92.95	94.89	95.09
19	10	3-Year	Low	High	–0.08	3.37	**10.33**	95.21	**74.16**	**38.39**
20	20	3-Year	Low	High	0.38	–0.14	–0.71	94.73	**85.94**	**52.36**
21	30	3-Year	Low	High	0.45	0.55	1.20	96.44	93.66	**75.49**
22	10	5-Year	Low	High	0.25	–0.47	**–8.03**	**87.07**	94.29	**29.85**
23	20	5-Year	Low	High	–0.17	–1.26	2.20	93.00	96.00	**37.44**
24	30	5-Year	Low	High	0.34	–0.01	1.80	93.63	95.07	**63.82**
25	10	10-Year	Low	High	**–7.14**	–4.38	–2.46	**80.17**	**92.28**	94.54
26	20	10-Year	Low	High	–0.13	–0.04	–0.13	**91.95**	95.31	95.27
27	30	10-Year	Low	High	0.07	–0.32	–0.14	**92.39**	93.21	94.27
28	10	3-Year	High	High	0.43	0.09	**–5.67**	94.65	**70.43**	**37.71**
29	20	3-Year	High	High	–1.14	–0.08	–1.15	94.17	**83.77**	**51.37**
30	30	3-Year	High	High	–0.96	–1.42	–2.30	95.42	93.39	**76.44**
31	10	5-Year	High	High	–1.38	–2.97	3.24	**89.11**	95.62	**26.29**
32	20	5-Year	High	High	0.79	1.00	–2.00	92.93	94.79	**37.42**
33	30	5-Year	High	High	0.36	0.38	1.25	93.74	94.42	**61.98**
34	10	10-Year	High	High	–1.98	–0.31	0.91	**81.73**	**92.33**	95.31
35	20	10-Year	High	High	–0.15	0.41	0.32	92.52	95.58	93.87
36	30	10-Year	High	High	0.23	0.44	0.68	94.39	94.75	94.79

Note. Items in bold indicate results outside of the acceptable boundaries for relative absolute bias or coverage. Shaded table cells indicate conditions where there is a match between the cohort selection mechanism in the data generating and estimating models. Con. = study condition; CSM = cohort selection mechanism; DGM = data generating model; Var. = variability.

overestimation as high as 10%.

Coverage rates for the level-1 curvilinear age parameter were largely driven by the cohort selection mechanism. Across all study conditions, when the cohort selection used in the data generating models and data estimating models matched, the curvilinear age effect was appropriately covered. Across all study conditions where the 10-year estimating model did not match the cohort selection mechanism in the data generating model, the curvilinear age parameter was under-covered (range = 26.29% to 91.49%), the magnitude of which increased as the number of repeated cross-sections decreased. There were many conditions in which the 3-year and 5-year estimating models could acceptably cover the curvilinear age parameter without needing to match the cohort selection mechanism in the data generating model (for example, see conditions 3 and 18). The general exception to this trend tended to occur when the data were generated using 10 repeated cross-sections. In these scenarios, the curvilinear age parameter was under-covered (range = 70.43% to 92.33%).

Coefficient of the Level-1 Predictor, β_3

Relative absolute bias and coverage rates of the level-1 predictor coefficient, β_3, are displayed in Table 9. Across all study conditions, the coefficient of the level-1 predictor was estimated without bias. Additionally, across all study conditions there was no evidence of under-coverage or over-coverage of the 95% confidence interval. Accordingly, the coefficient of the level-1 predictor was not impacted by any of the manipulated study conditions.

Coefficient of the Level-2 Cohort Predictor, γ_1

Relative absolute bias and coverage rates of the coefficient of the level-2 cohort predictor, γ_1, are displayed in Table 10. Relative absolute bias for the cohort predictor was most notably determined by the cohort selection mechanism. Across all study conditions where the 3-year

Table 9. Relative Absolute Bias (RAB) and Coverage Rates of the Coefficient of the Level-1 Predictor, β_3

Manipulated Condition					RAB by CSM in Estimating Model			Coverage Rates by CSM in Estimating Model		
Con.	Years	CSM in DGM	Period Var.	Cohort Var.	3-Year	5-Year	10-Year	3-Year	5-Year	10-Year
1	10	3-Year	Low	Low	0.92	0.79	0.68	95.42	95.34	95.07
2	20	3-Year	Low	Low	0.09	0.18	0.13	96.03	95.38	95.34
3	30	3-Year	Low	Low	0.16	0.21	0.65	93.88	94.28	94.18
4	10	5-Year	Low	Low	1.35	1.59	1.01	95.47	95.03	94.95
5	20	5-Year	Low	Low	–0.12	–0.17	–0.42	94.82	95.42	95.44
6	30	5-Year	Low	Low	–0.40	–0.59	–0.81	95.08	94.74	95.05
7	10	10-Year	Low	Low	0.69	0.35	0.02	94.56	94.51	94.86
8	20	10-Year	Low	Low	0.34	0.12	0.30	95.24	95.16	95.02
9	30	10-Year	Low	Low	0.11	0.29	0.36	94.31	93.92	94.43
10	10	3-Year	High	Low	1.33	1.46	1.85	94.20	94.22	94.49
11	20	3-Year	High	Low	–0.31	–0.44	–0.92	95.87	96.41	96.23
12	30	3-Year	High	Low	–0.33	–0.41	–0.36	95.09	95.48	95.47
13	10	5-Year	High	Low	0.16	0.45	0.14	95.65	95.18	95.06
14	20	5-Year	High	Low	–0.05	–0.21	0.04	95.32	95.88	94.43
15	30	5-Year	High	Low	–0.80	–0.83	–0.97	95.41	95.08	94.61
16	10	10-Year	High	Low	0.75	0.75	0.57	95.55	95.64	95.52
17	20	10-Year	High	Low	–0.05	0.06	–0.36	94.22	94.40	94.24
18	30	10-Year	High	Low	0.02	–0.09	–0.10	92.95	92.90	93.06
19	10	3-Year	Low	High	1.13	1.12	0.81	95.10	95.06	94.33
20	20	3-Year	Low	High	–0.42	–0.67	–0.48	95.05	94.99	94.62
21	30	3-Year	Low	High	0.77	0.72	0.73	94.56	94.30	93.98
22	10	5-Year	Low	High	0.38	–0.08	0.04	95.15	95.35	95.59
23	20	5-Year	Low	High	0.39	0.23	0.14	95.44	95.16	95.37
24	30	5-Year	Low	High	0.00	–0.04	–0.14	94.78	94.76	96.00
25	10	10-Year	Low	High	–0.05	–0.10	–0.17	94.24	94.29	94.65
26	20	10-Year	Low	High	0.10	–0.42	0.19	94.70	94.78	94.84
27	30	10-Year	Low	High	0.51	0.34	0.33	94.99	95.09	95.44
28	10	3-Year	High	High	–0.08	–0.48	0.30	93.58	93.77	94.49
29	20	3-Year	High	High	0.08	0.25	0.40	96.12	96.08	95.90
30	30	3-Year	High	High	0.06	–0.04	–0.27	94.14	93.82	93.22
31	10	5-Year	High	High	0.26	–0.06	–0.01	95.20	94.56	95.07
32	20	5-Year	High	High	–0.03	–0.07	0.11	93.35	93.73	94.04
33	30	5-Year	High	High	0.29	–0.10	0.15	93.96	94.21	94.43
34	10	10-Year	High	High	–2.05	–2.14	–1.82	95.94	95.95	96.14
35	20	10-Year	High	High	0.02	0.26	0.05	96.10	96.11	96.02
36	30	10-Year	High	High	–0.06	–0.11	0.02	96.30	95.80	95.44

Note. Items in bold indicate results outside of the acceptable boundaries for relative absolute bias or coverage. Shaded table cells indicate conditions where there is a match between the cohort selection mechanism in the data generating and estimating models. Con. = study condition; CSM = cohort selection mechanism; DGM = data generating model; Var. = variability.

Table 10. Relative Absolute Bias (RAB) and Coverage Rates of the Coefficient of the Level-2 Cohort Predictor, γ_1

Manipulated Condition					RAB by CSM in Estimating Model			Coverage Rates by CSM in Estimating Model		
Con.	Years	CSM in DGM	Period Var.	Cohort Var.	3-Year	5-Year	10-Year	3-Year	5-Year	10-Year
1	10	3-Year	Low	Low	1.00	**–25.89**	**–40.55**	94.14	**91.10**	**86.07**
2	20	3-Year	Low	Low	0.46	**–32.87**	–6.66	94.95	**90.00**	**87.92**
3	30	3-Year	Low	Low	–2.07	**–29.12**	**–25.55**	95.04	**91.68**	**89.01**
4	10	5-Year	Low	Low	**–54.94**	**9.76**	**23.14**	**82.30**	93.44	**89.34**
5	20	5-Year	Low	Low	**–56.11**	–3.74	–8.14	**82.40**	94.99	**85.89**
6	30	5-Year	Low	Low	**–59.62**	0.84	–8.29	**82.53**	94.11	**86.15**
7	10	10-Year	Low	Low	**–82.29**	**–45.47**	4.48	**66.63**	**89.65**	**90.56**
8	20	10-Year	Low	Low	**–77.84**	**–63.97**	0.50	**69.87**	**84.42**	93.11
9	30	10-Year	Low	Low	**–78.08**	**–55.19**	**9.01**	**70.60**	**86.69**	**92.33**
10	10	3-Year	High	Low	-3.08	**–31.49**	**–33.36**	95.15	**90.59**	**87.89**
11	20	3-Year	High	Low	-1.36	**–36.40**	–9.13	94.29	**89.43**	**87.25**
12	30	3-Year	High	Low	-0.76	**–28.63**	**–21.74**	**92.20**	**90.75**	**86.63**
13	10	5-Year	High	Low	**–55.93**	2.84	1.59	**81.21**	93.89	**88.78**
14	20	5-Year	High	Low	**–53.90**	2.86	**6.41**	**82.00**	92.60	**83.74**
15	30	5-Year	High	Low	**–57.13**	0.99	3.12	**81.11**	93.93	**87.21**
16	10	10-Year	High	Low	**–81.10**	**–45.41**	**7.77**	**64.41**	**89.10**	93.18
17	20	10-Year	High	Low	**–77.46**	**–62.51**	0.33	**72.27**	**85.62**	92.72
18	30	10-Year	High	Low	**–80.52**	**–60.44**	**–7.40**	**67.19**	**86.12**	93.81
19	10	3-Year	Low	High	–4.97	**–29.00**	**–37.36**	94.12	**90.02**	**82.99**
20	20	3-Year	Low	High	–2.11	**–49.47**	**6.73**	93.86	**85.94**	**79.91**
21	30	3-Year	Low	High	0.38	**–27.30**	**16.86**	93.41	**91.24**	**82.60**
22	10	5-Year	Low	High	**–59.42**	–4.07	**–13.15**	**89.01**	93.87	**83.37**
23	20	5-Year	Low	High	**–62.51**	**5.57**	**39.31**	**90.24**	94.32	**79.74**
24	30	5-Year	Low	High	**–59.94**	0.66	**22.50**	**87.16**	94.23	**80.24**
25	10	10-Year	Low	High	**–76.83**	**–41.77**	–4.87	**83.69**	**91.64**	**91.19**
26	20	10-Year	Low	High	**–74.33**	**–77.87**	**27.05**	**85.28**	**87.22**	**91.62**
27	30	10-Year	Low	High	**–77.39**	**–67.83**	**–23.94**	**85.40**	**88.40**	93.31
28	10	3-Year	High	High	**5.25**	**–21.84**	**–22.84**	94.43	**88.81**	**82.58**
29	20	3-Year	High	High	1.35	**–53.24**	2.24	93.31	**88.67**	**79.95**
30	30	3-Year	High	High	1.04	**–43.80**	**–29.89**	95.21	**89.06**	**83.56**
31	10	5-Year	High	High	**–56.97**	**6.89**	**–10.75**	**89.01**	93.60	**82.26**
32	20	5-Year	High	High	**–62.77**	**6.48**	**–28.13**	**89.60**	94.79	**82.02**
33	30	5-Year	High	High	**–63.57**	–0.97	**–72.85**	**88.03**	**91.47**	**78.37**
34	10	10-Year	High	High	**–76.83**	**–38.19**	**–6.99**	**83.97**	**91.69**	**92.18**
35	20	10-Year	High	High	**–80.50**	**–58.91**	**–14.46**	**85.35**	**89.79**	93.12
36	30	10-Year	High	High	**–79.97**	**–67.07**	**–6.46**	**84.13**	**89.19**	92.62

Note. Items in bold indicate results outside of the acceptable boundaries for relative absolute bias or coverage. Shaded table cells indicate conditions where there is a match between the cohort selection mechanism in the data generating and estimating models. Con. = study condition; CSM = cohort selection mechanism; DGM = data generating model; Var. = variability.

estimating model matched the data generating model, the cohort variable was recovered without bias. Trends for the 5-year and 10-year estimating models were less consistent. For conditions where the 5-year estimating model matched the data generating model (conditions 4-6, 13-15, 22-24, and 31-33), the model consistently recovered the cohort predictor when the data were generated with 30 repeated cross-sections. However, there were two conditions where the parameter was moderately overestimated when the data were generated using 10 repeated cross-sections (conditions 4, with 9.76% overestimation, and 31, with 6.89% overestimation), and two conditions where the parameter was moderately overestimated when the data were generated using 20 repeated cross-sections (conditions 23, with 5.57% overestimation, and 32, with 6.48% overestimation).

For conditions where the 10-year estimating model matched the data generating model (conditions 7-9, 16-18, 25-27, and 34-36), there were many scenarios where the cohort predictor was recovered with bias. When high variability was placed on both the period and cohort trends (conditions 34, 35, and 36), the cohort predictor was always underestimated, ranging from 6.99% to 14.46% underestimation. The 10-year estimating model also exhibited negative parameter bias in condition 18 (−7.40%) and condition 27 (−23.94%). Positive parameter bias was found in conditions 9, 16, and 26, where the cohort predictor variable was overestimated by 9.01%, 7.77%, and 27.05%, respectively. There was no readily apparent trend among the manipulated study conditions that explained these scenarios.

One interesting finding was that, generally, when the cohort selection mechanism between the data generating model and data estimating model did not match, the resulting cohort predictor was substantially biased. As an example, consider the dispersion plot pictured in Figure 3 for study conditions 14, 22, and 33. The data in all three conditions were generated using 5-

year cohort selection mechanisms. As visualized by the figure, when the estimating models utilized the 5-year cohort selection mechanism, the coefficient of the cohort predictor was recovered without bias. When the 3-year cohort selection mechanism was utilized in the estimating model, the level-2 cohort parameter was consistently underestimated by more than 50%, with very little variability in the estimated parameters. However, the estimates of the cohort parameter obtained using the 10-year estimating model exhibited more variability and resulted in the effect being underestimated as much as 73% (condition 33) and overestimated upwards of 6% (condition 14).

Coverage rates for the coefficient of the level-2 cohort predictor followed the same general pattern discovered for relative absolute bias. Generally, when the cohort selection mechanisms matched between the data generating and data estimating models, the cohort parameter was appropriately covered. The main exception to this finding was the 10-year estimating model, which had five conditions where the cohort selection mechanisms matched but the cohort parameter was slightly under-covered (conditions 7, 9, 25, 26, and 34, coverage range = 90.56% to 92.33%). Among these scenarios there was no readily apparent theme among the manipulated conditions that explained the under-coverage. Interestingly, across all study conditions, when the cohort selection mechanism in the data generating and data estimating models did not match, the cohort-level parameter was always under-covered, ranging from 64.41% under-coverage to 91.69% under-coverage. These results are displayed visually in a zipper plot in Figure 4 for study conditions 13, 22, and 31, all of which were generated using 10 repeated cross-sections and 5-year cohorts in the data generating model.

Coefficient of the Level-2 Period Predictor, γ_2

Relative absolute bias and coverage rates of the level-2 period predictor coefficient, γ_2,

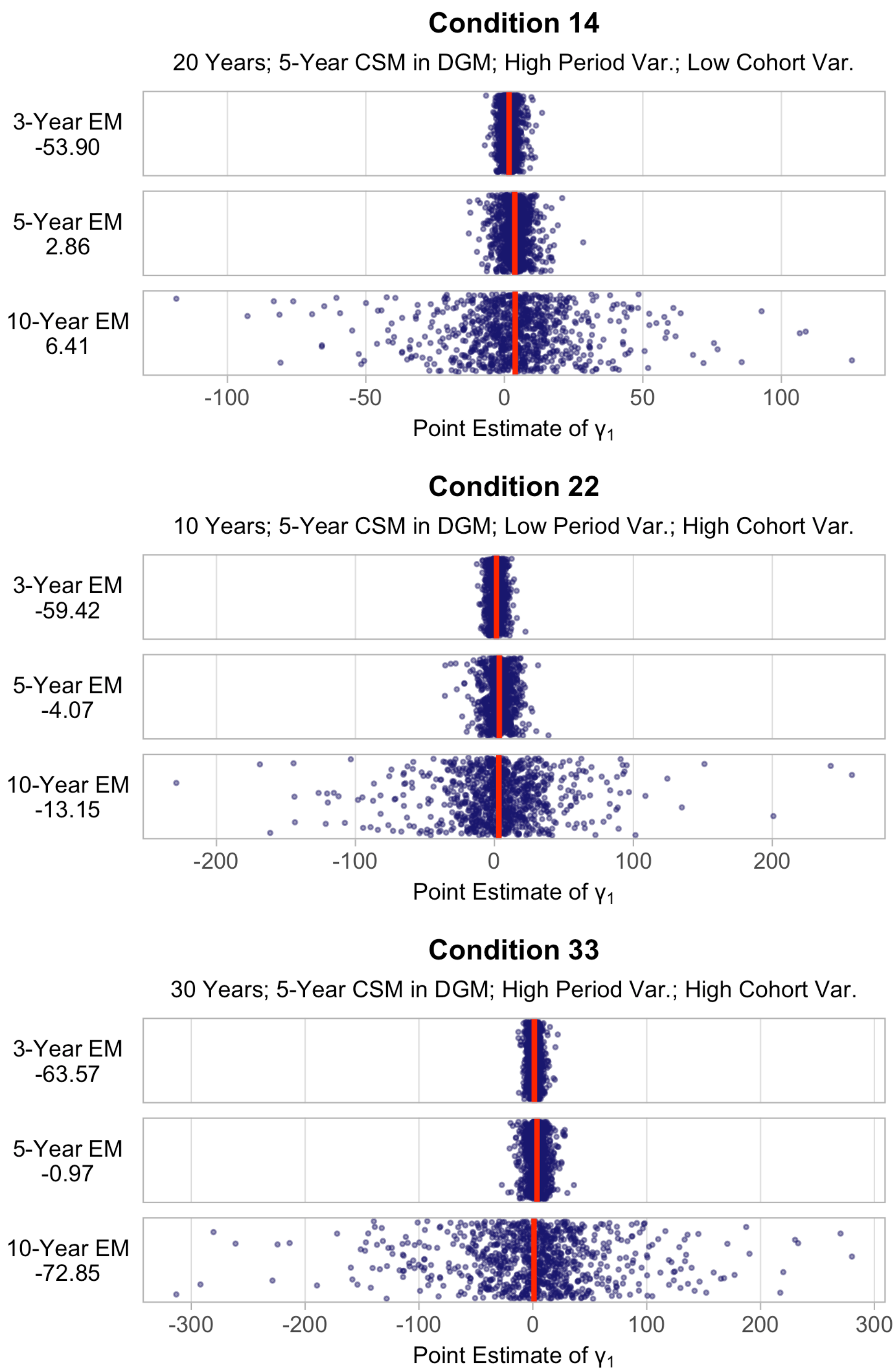

Figure 3. Dispersion Plot of the Coefficient of the Level-2 Cohort Predictor, γ_1, for Three Study Conditions

Note. Blue points represent the estimate of the cohort parameter for each successfully converged model. Red vertical lines represent the mean of the estimates among each cohort selection mechanism. Values under the estimating model label indicate the value of the relative absolute bias. CSM = cohort selection mechanism; DGM = data generating model; Var. = variability; EM = estimating model.

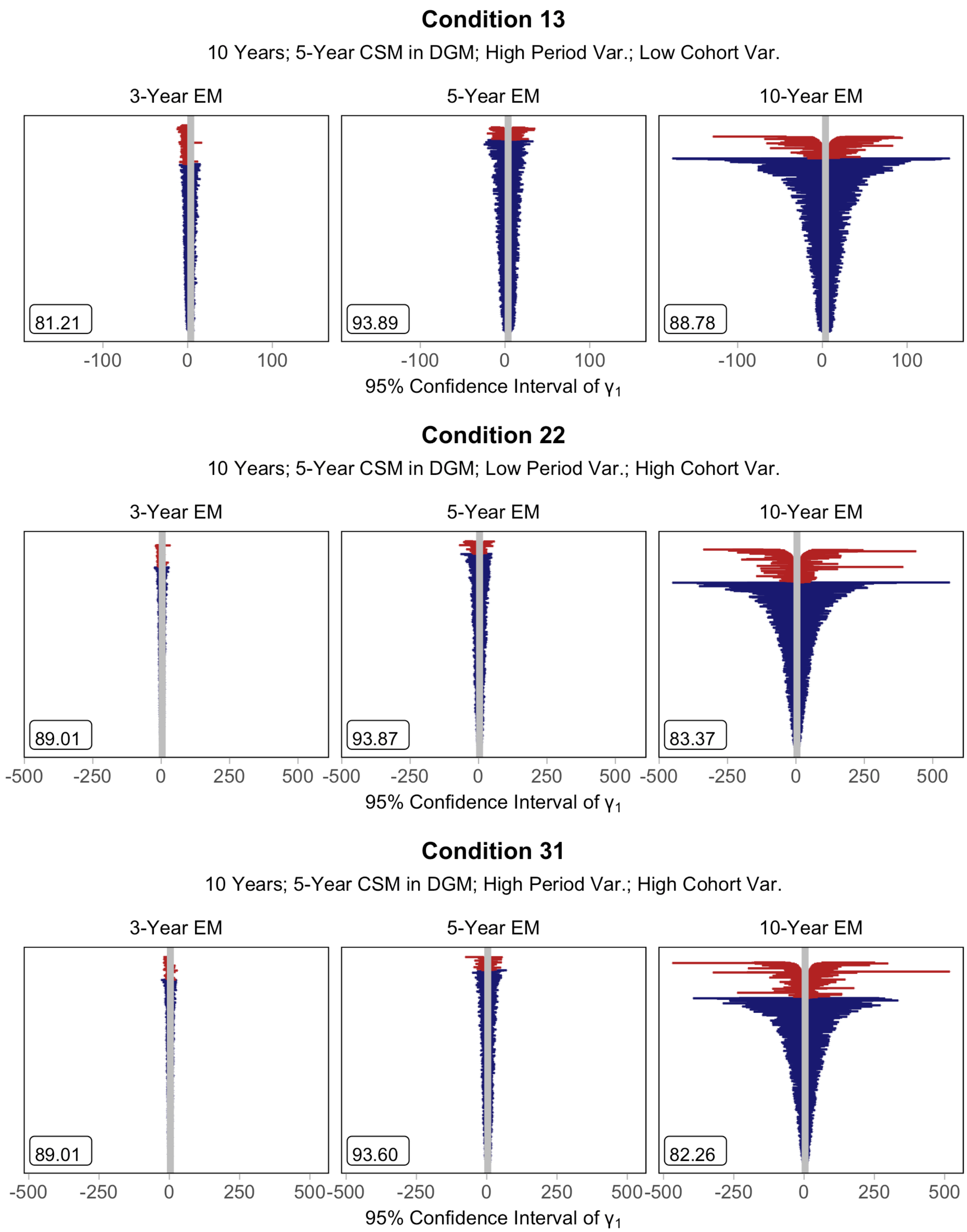

Figure 4. Zipper Plot of Coefficient of the Level-2 Cohort Predictor, γ_1, for Three Study Conditions

Note. CSM = cohort selection mechanism; DGM = data generating model; Var. = variability; EM = estimating model.

are displayed in Table 11. Relative absolute bias was largely influenced by the number of repeated cross-sections used to generate the data. Biased estimates of the period predictor's coefficient most commonly occurred when only 10 repeated cross-sections were used in the analysis, resulting in underestimation as much as 26.66% and overestimation as much as 30.99% of the effect, even when the cohort selection mechanism in the data estimating models and data generating models matched. Generally, for the datasets with 20 and 30 years of data, the estimating model was able to successfully recover the period-level parameter without bias. An exception to this pattern occurred in study conditions 28 through 36, where both the period and cohorts were generated with high variability. In these scenarios, the coefficient of the period-level predictor was recovered with relatively high bias, with the coefficient being underestimated by as much as 21.01% (condition 29). Notably, within these conditions the degree of bias improved as the number of repeated cross-sections increased. These findings are displayed visually in Figure 5.

Coverage rates of the level-2 period parameter were also predominately determined by the number of repeated cross-sections used to generate the data. Generally, for the datasets containing 20 and 30 years of data, the coefficient of the period predictor was acceptably covered regardless of the cohort selection mechanism and the amount of variability placed on the period and cohort clusters. However, in all of the conditions using 10 repeated cross-sections of data, the period predictor's coefficient was under-covered. Notably, the amount of under-coverage was small, generally falling between one to two percentage points outside of the acceptable range (the most severe under-coverage was 89.64%, occurring in study condition 22). There were two study conditions using 20 repeated cross-sections of data where the period predictor's coefficient was under-covered (conditions 5 and 20). Importantly, under-coverage only occurred in one of the three estimating models in these conditions, and the coverage rate was just outside

Table 11. Relative Absolute Bias (RAB) and Coverage Rates of the Coefficient of the Level-2 Period Predictor, γ_2

Manipulated Condition					RAB by CSM in Estimating Model			Coverage Rates by CSM in Estimating Model		
Con.	Years	CSM in DGM	Period Var.	Cohort Var.	3-Year	5-Year	10-Year	3-Year	5-Year	10-Year
1	10	3-Year	Low	Low	**9.08**	**9.49**	**7.44**	**90.95**	**90.25**	**90.68**
2	20	3-Year	Low	Low	**5.09**	3.66	3.87	94.41	94.62	94.12
3	30	3-Year	Low	Low	4.78	4.61	**5.43**	94.30	94.28	94.40
4	10	5-Year	Low	Low	**18.48**	**17.98**	**15.85**	**91.57**	**91.01**	**90.44**
5	20	5-Year	Low	Low	1.03	3.48	3.78	**92.44**	92.65	92.67
6	30	5-Year	Low	Low	–1.81	–1.51	–2.78	93.83	94.21	94.84
7	10	10-Year	Low	Low	0.01	1.82	0.23	**91.53**	**92.29**	**91.40**
8	20	10-Year	Low	Low	–1.49	–1.31	–0.80	92.71	93.16	93.01
9	30	10-Year	Low	Low	–2.25	–2.51	–2.96	94.94	94.76	94.54
10	10	3-Year	High	Low	**–8.01**	**–8.47**	**–13.32**	**91.04**	**91.66**	**91.08**
11	20	3-Year	High	Low	0.38	0.72	1.95	93.02	93.02	93.02
12	30	3-Year	High	Low	–0.61	**–5.20**	–1.79	93.80	93.80	93.26
13	10	5-Year	High	Low	**–15.67**	**–17.40**	**–24.84**	**91.30**	**91.10**	**91.69**
14	20	5-Year	High	Low	**–5.14**	**–6.06**	–0.69	93.86	93.87	93.76
15	30	5-Year	High	Low	4.22	**6.77**	3.57	93.84	93.61	93.94
16	10	10-Year	High	Low	**20.31**	**25.87**	**16.82**	**90.68**	**90.86**	**90.09**
17	20	10-Year	High	Low	1.90	1.24	**–5.03**	93.28	93.55	93.59
18	30	10-Year	High	Low	**–5.83**	–1.28	**–5.26**	92.53	92.80	92.74
19	10	3-Year	Low	High	**6.60**	**8.77**	**9.48**	**91.08**	**91.07**	**90.29**
20	20	3-Year	Low	High	–2.69	**–5.06**	–1.06	92.79	**92.44**	92.54
21	30	3-Year	Low	High	3.24	3.12	–2.46	93.62	93.45	93.98
22	10	5-Year	Low	High	**–15.32**	**–14.11**	**–12.86**	**90.09**	**89.64**	**90.53**
23	20	5-Year	Low	High	0.12	–1.74	**5.13**	94.06	94.21	93.94
24	30	5-Year	Low	High	2.57	1.04	**7.41**	93.63	93.50	93.67
25	10	10-Year	Low	High	**5.71**	**5.51**	2.63	**90.30**	**90.79**	**91.19**
26	20	10-Year	Low	High	0.06	–1.72	0.50	93.11	92.97	93.77
27	30	10-Year	Low	High	2.38	2.77	4.27	93.64	93.94	92.78
28	10	3-Year	High	High	**–19.48**	**–26.66**	**–25.43**	**90.47**	**91.02**	**91.07**
29	20	3-Year	High	High	**–15.50**	**–21.01**	**–24.32**	93.85	93.90	94.53
30	30	3-Year	High	High	**–5.33**	**–9.02**	–3.58	92.55	93.28	93.56
31	10	5-Year	High	High	**8.14**	**14.39**	**9.01**	**91.46**	**92.00**	**92.11**
32	20	5-Year	High	High	1.09	–0.39	**–10.58**	93.35	93.73	92.92
33	30	5-Year	High	High	2.99	1.30	**–7.57**	93.74	94.42	93.76
34	10	10-Year	High	High	**25.98**	**28.24**	**30.99**	**90.17**	**89.99**	**90.09**
35	20	10-Year	High	High	**–5.08**	**–5.16**	–2.97	93.78	93.58	93.23
36	30	10-Year	High	High	**–5.71**	**–5.86**	0.95	93.97	93.70	93.92

Note. Items in bold indicate results outside of the acceptable boundaries for relative absolute bias or coverage. Shaded table cells indicate conditions where there is a match between the cohort selection mechanism in the data generating and estimating models. Con. = study condition; CSM = cohort selection mechanism; DGM = data generating model; Var. = variability.

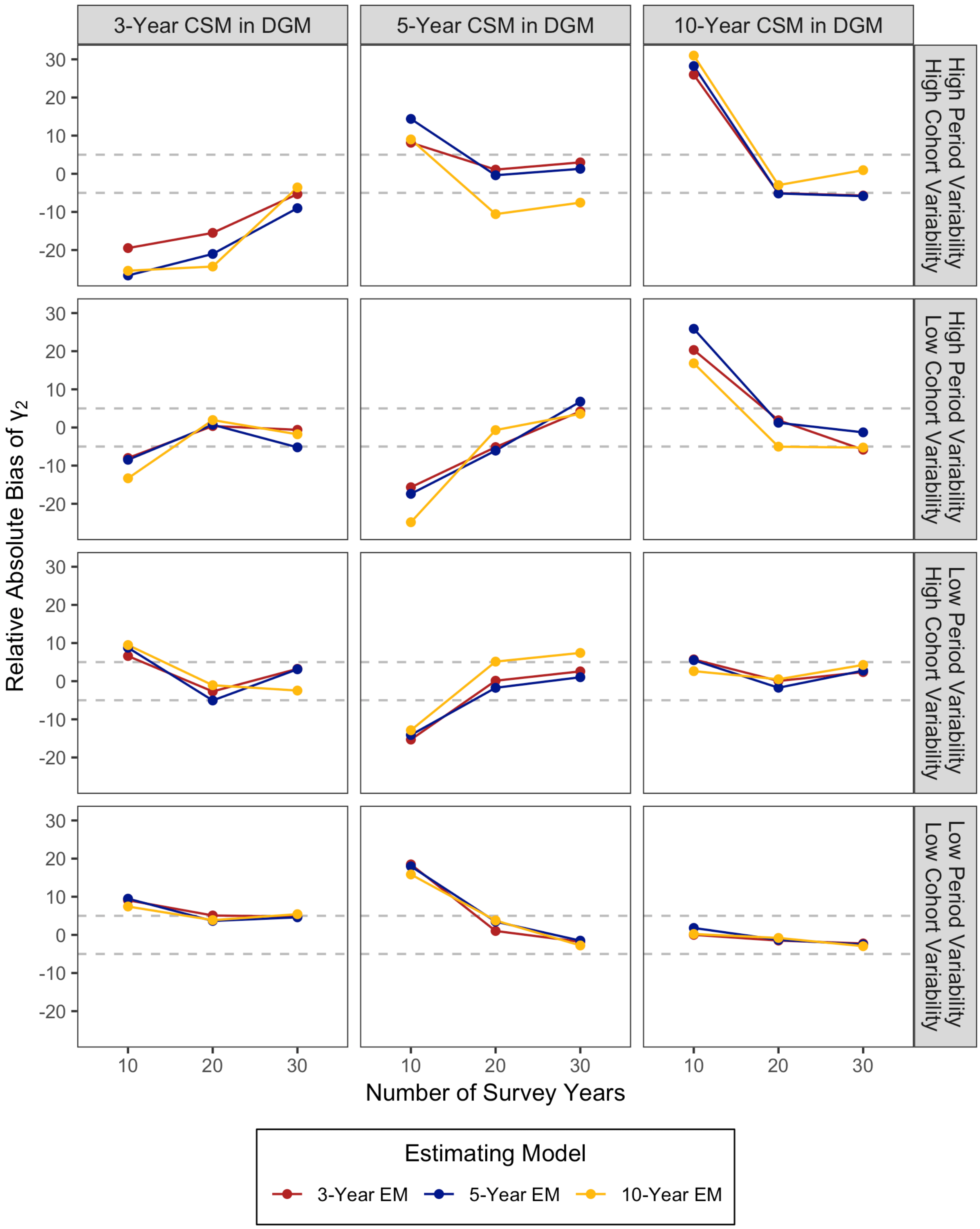

Figure 5. Relative Absolute Bias of the Coefficient of the Level-2 Period Predictor, γ_2

Note. Grey dashed lines represent acceptable boundaries of relative absolute bias (−5 to 5). CSM = cohort selection mechanism; DGM = data generating model; EM = estimating model.

the acceptable coverage range (92.44% in both conditions).

Relative Absolute Bias of Variance Components Estimates

Level-1 Variance Component, σ^2

Relative absolute bias of the level-1 individual variance component, σ^2, is displayed in Table 12. There were only eight conditions where the level-1 variance component was recovered with moderate bias. The common factor across these conditions was that the cohorts were generated with high variability and the cohort selection mechanism in the data generating model did not match the data estimating model. In these conditions, the level-1 variance component was overestimated between 5.71% to 8.72%. Otherwise, recovery of the level-1 variance component did not appear to be influenced by any of the other manipulated study conditions.

Level-2 Cohort Variance Component, τ_u

Relative absolute bias of the level-2 cohort variance component, τ_u, is displayed in Table 13. Successful recovery of the cohort variance component was determined by the cohort selection mechanism used in the estimating model matching the cohort selection mechanism used to generate the data. Across all study conditions, when the cohort selection mechanisms matched, the cohort variance component was estimated without bias. One interesting finding is that the magnitude of the relative absolute bias appeared to be influenced by the degree of mismatch between the cohort selection mechanisms. Most notably, when the data were generated using 5-year cohorts, the 3-year estimating models always underestimated the cohort variance component (between 16.29% to 25.38% under-estimation). At the same time, the 10-year estimating models drastically overestimated the cohort variance component, especially when the cohorts were generated with high variability. Indeed, when the cohorts were generated with low variability, the cohort variance component was overestimated between 210.71% (condition 4) and 315.85% (condition 15). Over-estimation of the cohort variance component increased to as much as

Table 12. Relative Absolute Bias of Level-1 Variance Component, σ^2

Manipulated Condition					CSM in Estimating Model		
Condition	Years	CSM DGM	Period Variability	Cohort Variability	3-Year	5-Year	10-Year
1	10	3-Year	Low	Low	0.00	1.10	1.50
2	20	3-Year	Low	Low	–0.03	1.06	1.46
3	30	3-Year	Low	Low	0.04	1.12	1.54
4	10	5-Year	Low	Low	0.39	0.01	0.91
5	20	5-Year	Low	Low	0.40	0.02	0.92
6	30	5-Year	Low	Low	0.34	–0.03	0.90
7	10	10-Year	Low	Low	0.20	0.00	0.02
8	20	10-Year	Low	Low	0.18	0.00	0.00
9	30	10-Year	Low	Low	0.19	0.01	0.00
10	10	3-Year	High	Low	0.05	1.22	1.64
11	20	3-Year	High	Low	0.01	1.17	1.63
12	30	3-Year	High	Low	0.03	1.19	1.63
13	10	5-Year	High	Low	0.41	–0.01	0.97
14	20	5-Year	High	Low	0.36	–0.04	0.95
15	30	5-Year	High	Low	0.43	0.01	1.03
16	10	10-Year	High	Low	0.17	–0.02	–0.03
17	20	10-Year	High	Low	0.21	–0.01	–0.01
18	30	10-Year	High	Low	0.25	0.04	0.03
19	10	3-Year	Low	High	0.03	**5.83**	**7.90**
20	20	3-Year	Low	High	0.02	**5.71**	**7.91**
21	30	3-Year	Low	High	–0.01	**5.72**	**7.99**
22	10	5-Year	Low	High	1.89	–0.09	4.47
23	20	5-Year	Low	High	2.05	–0.01	4.98
24	30	5-Year	Low	High	2.04	0.01	**5.04**
25	10	10-Year	Low	High	0.93	–0.07	–0.07
26	20	10-Year	Low	High	1.06	0.02	0.02
27	30	10-Year	Low	High	0.96	–0.04	–0.04
28	10	3-Year	High	High	0.02	**6.17**	**8.49**
29	20	3-Year	High	High	0.00	**6.17**	**8.70**
30	30	3-Year	High	High	0.00	**6.23**	**8.72**
31	10	5-Year	High	High	2.22	0.03	**5.12**
32	20	5-Year	High	High	2.32	–0.02	**5.64**
33	30	5-Year	High	High	2.19	0.01	**5.46**
34	10	10-Year	High	High	1.00	–0.05	–0.06
35	20	10-Year	High	High	1.13	0.00	0.00
36	30	10-Year	High	High	1.10	0.01	0.01

Note. Items in bold indicate results outside of the acceptable boundaries for relative absolute bias. Shaded table cells indicate conditions where there is a match between the cohort selection mechanism in the data generating and estimating models. CSM DGM = cohort selection mechanism used in the data generating model.

Table 13. Relative Absolute Bias of Level-2 Cohort Variance Component, τ_u

Manipulated Condition					CSM in Estimating Model		
Condition	Years	CSM DGM	Period Variability	Cohort Variability	3-Year	5-Year	10-Year
1	10	3-Year	Low	Low	0.92	**–49.26**	–2.73
2	20	3-Year	Low	Low	–0.38	**–49.85**	**10.95**
3	30	3-Year	Low	Low	0.43	**–49.80**	**–8.29**
4	10	5-Year	Low	Low	**–22.81**	1.50	**210.71**
5	20	5-Year	Low	Low	**–21.82**	–0.25	**255.75**
6	30	5-Year	Low	Low	**–21.66**	–0.39	**269.09**
7	10	10-Year	Low	Low	**–33.95**	**–16.46**	0.88
8	20	10-Year	Low	Low	**–28.20**	**–14.08**	–2.70
9	30	10-Year	Low	Low	**–25.03**	**–11.24**	–2.78
10	10	3-Year	High	Low	–0.88	**–49.22**	**22.44**
11	20	3-Year	High	Low	–0.54	**–49.50**	**31.51**
12	30	3-Year	High	Low	–0.69	**–49.95**	**32.32**
13	10	5-Year	High	Low	**–24.76**	–0.33	**248.14**
14	20	5-Year	High	Low	**–22.65**	–0.03	**337.24**
15	30	5-Year	High	Low	**–21.53**	–0.01	**315.85**
16	10	10-Year	High	Low	**–35.68**	**–18.37**	–4.51
17	20	10-Year	High	Low	**–25.42**	**–9.99**	2.95
18	30	10-Year	High	Low	**–21.99**	**–7.52**	1.75
19	10	3-Year	Low	High	0.34	**24.31**	**97.86**
20	20	3-Year	Low	High	–0.18	**56.64**	**114.99**
21	30	3-Year	Low	High	–0.82	**18.16**	**115.62**
22	10	5-Year	Low	High	**–25.38**	–0.83	**355.61**
23	20	5-Year	Low	High	**–20.79**	0.57	**414.81**
24	30	5-Year	Low	High	**–20.27**	0.62	**509.42**
25	10	10-Year	Low	High	**–25.88**	**–9.34**	2.31
26	20	10-Year	Low	High	**–21.43**	**–6.22**	1.99
27	30	10-Year	Low	High	**–20.43**	**–6.98**	0.83
28	10	3-Year	High	High	–0.99	**48.78**	**98.39**
29	20	3-Year	High	High	0.12	**90.26**	**117.30**
30	30	3-Year	High	High	–0.55	**96.90**	**142.10**
31	10	5-Year	High	High	**–23.43**	–0.98	**347.99**
32	20	5-Year	High	High	**–16.29**	3.07	**447.47**
33	30	5-Year	High	High	**–19.67**	–0.76	**531.50**
34	10	10-Year	High	High	**–30.39**	**–13.43**	0.27
35	20	10-Year	High	High	**–21.17**	**–9.59**	1.20
36	30	10-Year	High	High	**–19.91**	**–8.13**	0.72

Note. Items in bold indicate results outside of the acceptable boundaries for relative absolute bias. Shaded table cells indicate conditions where there is a match between the cohort selection mechanism in the data generating and estimating models. CSM DGM = cohort selection mechanism used in the data generating model.

531.50% (condition 33) when the cohorts were generated with high variability. This relationship is displayed visually using a dispersion plot in Figure 6 for three study conditions where the 10-year cohort selection mechanism used in the estimating model drastically overestimated the cohort variance component.

When moving in the opposite direction for the degree of mismatch between the cohort selection mechanism in the data estimating and data generating models, the same pattern occurs, but the magnitude of the biased parameter recovery is much less severe. Specifically, when 10-year cohorts were used to generate the data, the cohort variance component was always underestimated by the 3-year and 5-year estimating models, but the 3-year estimating models more greatly underestimated the cohort variance component compared to the 5-year models. In these scenarios, the 5-year estimating models underestimated the cohort variance component between 6.22% to 18.37%, while the 3-year estimating models underestimated between 20.43% to 35.68%.

Level-2 Period Variance Component, τ_v

Relative absolute bias of the level-2 period variance component, τ_v, is displayed in Table 14. Overestimation of the period variance component occurred when the cohort selection mechanism in the data generating models did not match the data estimating models, but only for the 5-year and 10-year estimating models (the 3-year estimating models exhibited no relative absolute bias across all study conditions). In the study conditions where the 5-year and 10-year estimating models resulted in overestimation of the period variance component, the degree of overestimation increased as the number of repeated cross-sections used to generate the data increased. For example, in study conditions 19 through 21, where 3-year cohorts were used to generate the data, the 5-year cohort selection mechanism overestimated the period variance

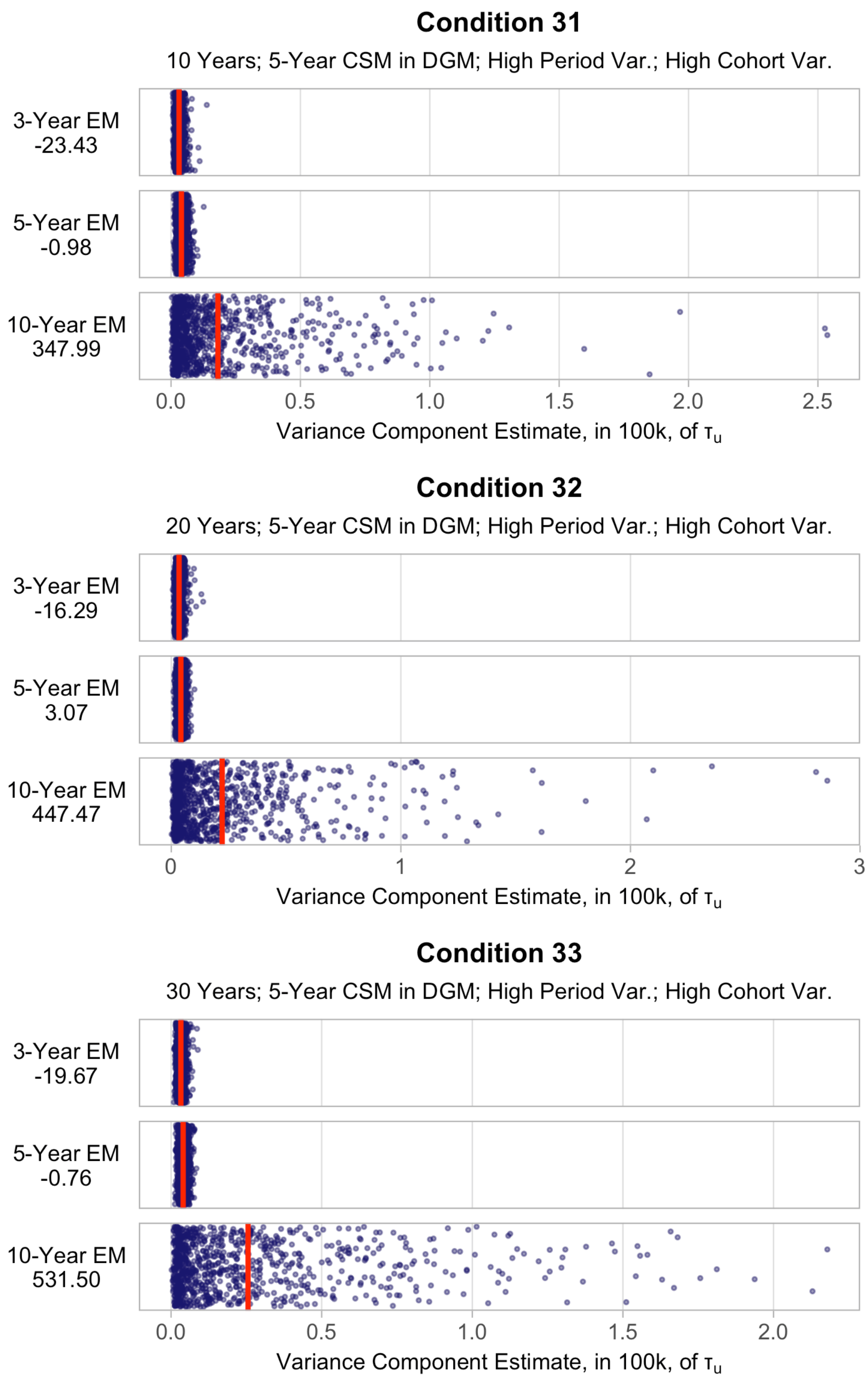

Figure 6. Dispersion Plot of Level-2 Cohort Variance Component, τ_u, for Three Study Conditions

Note. Blue points represent the estimate of the cohort parameter for each successfully converged model. Red vertical lines represent the mean of the estimates among each cohort selection mechanism. Values under the estimating model label indicate the value of the relative absolute bias. CSM = cohort selection mechanism; DGM = data generating model; Var. = variability; EM = estimating model.

component by 7.62%, 39.33%, and 57.38% as the number of repeated cross-sections increased from 10 to 20 to 30 years, respectively. The corresponding overestimation for the 10-year estimating model in the same study conditions was 20.26%, 69.30%, and 151.36%.

There were many more study conditions where the 10-year estimating models overestimated the period variance component compared to the 5-year estimating models, and generally the magnitude of the overestimation was higher for the 10-year estimating models. Specifically, for only five study conditions was the period variance component overestimated when 5-year cohorts were used in the estimating models (coverage range = 7.62% to 57.38%), while the 10-year estimating models overestimated the period variance component in 17 study conditions (coverage range = 5.86% to 443.95%). Overestimation of the period variance component using 10-year cohorts was particularly large when the cohorts were generated with high variability. For example, in study conditions 23, 24, and 33, the period variance component was overestimated by 174.12%, 443.95%, and 94.22%, respectively.

Table 14. Relative Absolute Bias of Level-2 Period Variance Component, τ_v

Manipulated Condition					CSM in Estimating Model		
Condition	Years	CSM DGM	Period Variability	Cohort Variability	3-Year	5-Year	10-Year
1	10	3-Year	Low	Low	–0.80	–0.65	0.67
2	20	3-Year	Low	Low	0.80	0.29	**5.53**
3	30	3-Year	Low	Low	0.17	0.91	**10.24**
4	10	5-Year	Low	Low	–0.78	–1.73	3.93
5	20	5-Year	Low	Low	–0.19	0.05	**20.31**
6	30	5-Year	Low	Low	–0.23	–0.79	**49.23**
7	10	10-Year	Low	Low	0.24	0.38	0.38
8	20	10-Year	Low	Low	1.30	1.05	0.47
9	30	10-Year	Low	Low	0.01	–0.56	–0.93
10	10	3-Year	High	Low	0.40	1.10	1.23
11	20	3-Year	High	Low	–1.89	–1.98	–0.31
12	30	3-Year	High	Low	0.60	0.71	3.47
13	10	5-Year	High	Low	–0.39	–0.62	1.65
14	20	5-Year	High	Low	0.09	0.13	**5.86**
15	30	5-Year	High	Low	0.18	–0.51	**10.40**
16	10	10-Year	High	Low	–1.41	–1.12	–1.34
17	20	10-Year	High	Low	–1.57	–2.00	–1.80
18	30	10-Year	High	Low	0.82	1.01	0.70
19	10	3-Year	Low	High	–1.15	**7.62**	**20.26**
20	20	3-Year	Low	High	1.10	**39.33**	**69.30**
21	30	3-Year	Low	High	–1.03	**57.38**	**151.36**
22	10	5-Year	Low	High	–1.50	–1.60	**42.26**
23	20	5-Year	Low	High	–1.43	–3.37	**174.12**
24	30	5-Year	Low	High	1.48	–0.82	**443.95**
25	10	10-Year	Low	High	0.60	–0.39	–0.37
26	20	10-Year	Low	High	1.36	–1.12	–1.75
27	30	10-Year	Low	High	4.22	1.57	0.71
28	10	3-Year	High	High	–0.87	1.17	3.01
29	20	3-Year	High	High	0.59	**10.63**	**13.80**
30	30	3-Year	High	High	0.38	**22.79**	**33.72**
31	10	5-Year	High	High	0.41	–0.16	**9.81**
32	20	5-Year	High	High	0.43	–0.01	**35.43**
33	30	5-Year	High	High	1.55	0.90	**94.22**
34	10	10-Year	High	High	2.39	1.51	1.07
35	20	10-Year	High	High	1.99	0.90	0.65
36	30	10-Year	High	High	1.34	–0.05	0.14

Note. Items in bold indicate results outside of the acceptable boundaries for relative absolute bias. Shaded table cells indicate conditions where there is a match between the cohort selection mechanism in the data generating and estimating models. CSM DGM = cohort selection mechanism used in the data generating model.

5 DISCUSSION

Proponents of life course theory contend that longitudinal analyses of social change should account for three important temporal forces: age, period, and cohort effects. However, simultaneous estimation of these three effects in the same model poses an identification problem because the three effects are linear derivates of one another, resulting in a perfectly specified model. For the past few decades, social scientists have utilized several different modeling techniques to overcome or work around the identification problem. One of the more recent solutions that has been particularly popular in the applied literature is the hierarchical-age-period-cohort cross-classified random effects model (HAPC-CCREM). Proposed by Yang and Land (2006), the HAPC-CCREM provides a unique solution to the identification problem by taking advantage of the multilevel modeling framework to specify age effects using individual-level data at level one, which are then cross-classified by period and cohort clusters at level two. To date, more than fifty studies have utilized the HAPC-CCREM to analyze social change in a variety of outcomes.

Despite the increasing popularity of the HAPC-CCREM in the applied literature, comparatively fewer methodological studies exist on the model. Utilizing a Monte Carlo simulation study, the purpose of this dissertation was to expand our methodological research on the HAPC-CCREM by examining several factors that may influence the accuracy of the estimates in the model. Furthermore, this dissertation also explored the performance of various model fit indices in identifying the underlying data structure used in the model. This chapter discusses the findings and implications of the Monte Carlo simulation study and is structured as follows. First, a summary of findings for each area investigated in the simulation study is provided, with specific attention devoted to explaining why the results may have occurred and how they relate to previous methodological research. Second, implications of the study findings are discussed, including implications for applied researchers, implications for administrators and funders of datasets used in HAPC-CCREM analyses, and implications for life course theory. Third, the chapter concludes with a discussion of important limitations of the study and highlights a variety of areas for future methodological research on the model.

The Influence of the Number of Repeated Cross-Sections

To date, only one previous study has examined the impact that the number of repeated cross-sections has on the accuracy of the model's estimates. Yang (2006) examined if the accuracy of the model estimates differed between datasets comprised of 5 repeated cross-sections compared to 15 repeated cross-sections, finding that in the smaller datasets the level-1 age effect, period variance component, and cohort variance component were all overestimated and the level-1 individual variance component was underestimated. For three reasons, this dissertation expanded on the Yang study. First, since the time of Yang's study, datasets used in HAPC-CCREM analyses have substantially grown, and while she found that larger datasets accurately recovered parameters in the model, this assumption remained untested in datasets with more than 15 years of data. Second, the findings of Yang's simulation imply that applied researchers should only use the HAPC-CCREM on datasets with at least 15 repeated cross-sections; however, a dataset of size between 5 and 15 years may accurately produce model estimates. Third, Yang's study did not examine the impact that the number of repeated cross-sections has on the coefficient estimate of a period-level predictor.

For these reasons, this dissertation examined the influence of the number of repeated cross-sections on the accuracy of parameter estimates in the HAPC-CCREM. Specifically, datasets with 10, 20, and 30 repeated cross-sections were examined. The number of repeated cross-sections most directly influenced the accuracy of the coefficient estimate of the level-2 period predictor. The findings of the simulation study indicated that the level-2 period predictor coefficient was underestimated by as much as 27%, overestimated by as much as 31%, and slightly under-covered (worst under-coverage = 90% coverage rate) when the datasets were generated using 10 repeated cross-sections. Otherwise, for datasets generated with 20 and 30 repeated cross-sections, the coefficient estimate of the level-2 period predictor was unaffected.

The number of repeated cross-sections partially impacted the following parameters. Coverage of the intercept was partially influenced; there were five conditions in which the intercept was slightly under-covered when 10 repeated cross-sections were used to generate the data (coverage range = 91.75% to 92.39%). Coverage of the level-1 linear age parameter was partially influenced as well. There were three conditions where the data were generated and estimated using 10-year cohorts, in which the level-1 linear age parameter was slightly under-covered (91.21% to 92.34%) for datasets generated with 10 repeated cross-sections. The level-1 curvilinear age parameter was partially influenced by the number of repeated cross-sections in 10-year estimating models that did not match the data generating models, for which datasets with 10 and 20 repeated cross-sections consistently exhibited lower coverage rates. Finally, the level-2 period variance component was partially influenced. Specifically, in conditions where the 5-year and 10-year cohort selection mechanisms used in the estimating models did not match the cohort selection mechanisms in the data generating models, overestimation of the period variance component appeared to increase as the number of repeated cross-sections increased.

The finding that the number of repeated cross-sections directly influences the accuracy of coefficient estimates of the level-2 period predictor and partially influences some of the other parameters in the HAPC-CCREM is an unsurprising finding. This is because the number of repeated cross-sections directly changes the number of level-2 period clusters used in the estimation of the model's parameters. As more years are available for analysis, the number of level-2 period clusters increases. Importantly, the number of clusters used in the estimation of a multilevel model is ultimately an issue of sample size, which has been of great concern to multilevel modelers. Multilevel models estimated using maximum-likelihood are asymptotically unbiased, meaning they perform well as their sample sizes approach infinity but perform comparatively worse with smaller samples (McNeish & Stapleton, 2016a). From a design perspective, it is generally recommended that in multilevel models the number of level-two clusters is at least fifty (Maas & Hox, 2002). However, previous simulation studies for a simple two-level model have indicated that for continuous outcomes the level-2 fixed effects can be accurately estimated with as few as 15 clusters, level-1 fixed effects can be accurately estimated with as few as five clusters, and level-1 and level-2 variance components can be estimated with as few as ten clusters (McNeish, 2014; McNeish & Harring, 2017; McNeish & Stapleton, 2016a, 2016b). Simulation studies of the cross-classified random effects model typically set the number of level-two clusters for one of the cross-classified factors at 20, 30, or 50 clusters and have found that with those numbers of clusters the model parameters could be accurately estimated (Kim et al., 2021; Luo & Kwok, 2009; Meyers & Beretvas, 2006; Ye & Daniel, 2017). Accordingly, the findings of this dissertation are in line with the previous simulation studies on the broader class of multilevel models.

The Influence of the Cohort Selection Mechanism

A small handful of simulation studies have previously examined the impact that the cohort selection mechanism has on the accuracy of the model estimates. These previous studies mostly focused on the ability of the model to accurately recover the age, period, and cohort trends when the cohorts used in the estimating model did not match the data generating model (Bell & Jones, 2014b, 2014c, 2015; Reither et al., 2015). Some of these simulation studies found that the model could not recover the age, period, and cohort trends, but the simulation conducted by Reither et al. (2015) indicated that these findings were erroneous because in the simulation design perfectly linear trends were specified in the data generating models, and that when more realistic, linear monotonic trends were specified the model could accurately recover the trends, even if the cohort groupings in the estimating model did not match the data generating model. This dissertation shifted attention away from the impact of the cohort selection mechanism on recovery of the age, period, and cohort trends, and instead examined accuracy of the model parameter estimates (intercept, age predictors, level-1 predictor, level-2 predictor, and variance components) given the cohort selection mechanism employed in the model. Specific attention was devoted to the impact of the cohort selection mechanism when the cohort groupings in the estimating model did and did not match those used in the data generating model. Motivated by common cohort groupings used in the applied research literature, groups of 3-year, 5-year, and 10-year cohorts were specifically analyzed.

Of the various factors examined in this simulation study, the cohort selection mechanism appeared to be the most influential in determining accuracy of the estimates in the HAPC-CCREM. While the cohort selection mechanism did not impact the accuracy of the coefficient estimates of the level-1 predictor or the level-2 period predictor, it did impact the other model parameters. Most notably, the cohort selection mechanism heavily impacted the accuracy of the level-1 linear age parameter, the level-2 cohort predictor coefficient, and the level-2 cohort variance component. For these three parameters, when the cohort selection mechanism in the estimating models did not match the data generating models the parameter was generally estimated with bias, and for the linear age parameter and cohort predictor the 95% confidence intervals of the estimate were generally under-covered. Moreover, for the cohort variance component, the magnitude of the parameter bias was driven by the degree of the mismatch between the cohort selection mechanism used in the estimating model versus the data generating model. However, when the cohort selection mechanism used in the estimating model matched the data generating model, the parameters were recovered without bias and the 95% confidence intervals (for the linear age parameter and cohort predictor) were appropriately covered. The findings are likely explainable by the direct connection between cohorts and these specific parameters. It makes sense for the cohort-level predictor coefficient and the cohort variance component to be directly influenced by the structure of the cohorts in the underlying dataset, and for these cohort-related variables to perform well when they match.

To a lesser extent, the cohort selection mechanism impacted the intercept, level-1 curvilinear age parameter, the level-1 variance component, and the level-2 period variance component. Relative absolute bias of the intercept was not influenced by the cohort selection mechanism, but when the cohort selection mechanism in the estimating models did not match the data generating models, the 95% confidence intervals of the intercept were consistently under-covered. A somewhat similar effect was found for the level-1 curvilinear age parameter. For the curvilinear age parameter, relative absolute bias was only apparent in four of the 36 study conditions, and in three of these four conditions, the models were estimated with 10-year cohort groupings that did not match the data generating model. For coverage of the curvilinear age parameter, many of the 3-year and 5-year estimating models that did not match the data generating model could appropriately cover the curvilinear age parameter, but 10-year estimating models that did not match exhibited significant under-coverage of the curvilinear age parameter. For the level-1 variance component, there were 10 study conditions where bias of this parameter was driven by the cohort selection mechanism in the estimating models not matching the data generating models, but these conditions were also all generated with high cohort variability. Finally, one interesting finding was that for the level-2 period variance component, the 3-year estimating models could accurately recover this parameter regardless of the cohort selection mechanism in the underlying data. However, 5-year and 10-year estimating models that did not match the data generating model consistently overestimated the period variance component, with the 10-year estimating models performing especially poor (overestimation ranged from 6% to 444%).

Taken together, the results of this simulation study indicate that the cohort selection mechanism can ultimately create model misspecification issues for estimation of the HAPC-CCREM. Model misspecification occurs when an estimated model fails to capture an important or meaningful aspect of the underlying data used by the model (Dennis et al., 2019). For example, in multilevel growth curve models model misspecification would occur when the underlying functional form of time is non-linear, but the model is estimated with a linear time effect (McCoach & Kaniskan, 2010; Singer & Willett, 2003); the model is mis-specified because the researcher failed to capture the non-linear time effect. As it relates to the HAPC-CCREM, the model misspecification occurs when the cohort selection mechanism used in the estimating model does not match the cohort structure of the underlying data set. For the HAPC-CCREM, model misspecification in this regard is likely so influential because membership in the level-2 cohort cluster is directly influenced by how the researcher decides to group individuals into the cluster, and this decision must be made for every individual in the dataset. While previous research indicates that the cohort grouping does not impact the ability of the model to recover the age, period, and cohort trends in the data (Reither et al., 2015), the findings of this dissertation indicate that failure to capture the cohort groupings underlying the data heavily impacts the accuracy of many of the parameters in the HAPC-CCREM.

The Influence of Variability in the Period and Cohort Effects

To date, none of the previous simulation studies on the HAPC-CCREM have examined the influence of the variability in the period and cohort effects on accuracy of the model's estimates. Therefore, this dissertation expanded on previous research by examining relative absolute bias and coverage rates of estimates in the HAPC-CCREM when the period effect exhibited low and high variability, as well as when the cohort effect exhibited low and high variability. Low variability was operationalized as 2% of the total variability in the outcome occurring for the period or cohort effect, and high variability was operationalized as 10% of the total variability in the outcome occurring for the period or cohort effect. The results of the simulation study indicated that variability in the period and cohort effects most clearly impacted the coefficient estimate of the level-2 period predictor. When both effects were generated with high variability, the period predictor coefficient was underestimated by as much as 21% in datasets with 10 repeated cross-sections. In datasets with 20 and 30 repeated cross-sections, the period predictor coefficient was underestimated by as much as 27% and overestimated by as much as 31% when variability in both the period and cohort effects was high. Additionally, the coefficient estimate of the level-2 cohort predictor was impacted by the period and cohort variability. Specifically, when the periods and cohorts were generated with high variability and a 10-year cohort selection mechanism, the cohort predictor coefficient was underestimated across all the estimating models, even when the cohort groupings in the estimating models matched the data generating models.

The intercept, level-1 curvilinear age parameter, and the level-1 predictor coefficient were not influenced by the period and cohort variability in any study conditions. Otherwise, the remaining parameters in the HAPC-CCREM were only influenced by the period and cohort variability when the cohort selection mechanism used in the estimating models did not match the data generating models. This trend was most prominent for the level-2 cohort variance component and the level-2 period variance component. When the estimating models with a 10-year cohort selection mechanism did not match the data generating models, overestimation of the cohort variance component increased as the cohort variability increased. Specifically, when the cohort variability was low, the cohort variance component was overestimated between 211% to 315%, which increased to as much as 532% when cohort variability was high. When the cohort selection mechanism used in the estimating models did not match the data generating models and the cohort variability was high, the period variance component was especially overestimated, ranging from 10% overestimation to 444% overestimation. The level-1 linear age parameter and the level-1 variance component were also impacted by period and cohort variability when the cohort selection mechanism in the estimating model did not match the data generating model, but comparatively less than the period and cohort variance component.

The findings of this simulation study that the coefficient estimates of the period and cohort-level predictors were impacted by variability in the period and cohort effects are interesting given the results of previous simulation studies of cross-classified random effects models. In terms of multilevel modeling, manipulating the variability of the period and cohort effects ultimately impacts the total level of variability at level one of the model and level two of the model. In the current study, the specific combinations of the manipulated period and cohort variability results in models with the total variability at level two ranging from 4% (periods and cohort both generated with low variability) to 20% (periods and cohorts both generated with high variability). Three previous simulation studies have examined the ability of a CCREM to recover parameters given the total variability at level two (Kim et al., 2021; Meyers & Beretvas, 2006; Ye & Daniel, 2017). All three studies evaluated the model with 5% and 15% of the total variability set at level two, and all three studies found that none of the level-1 or level-2 predictor variables' coefficients and variance components were impacted by the amount of variability occurring at level two.

One important difference in the design of the current simulation study and previous research on the CCREM is the number of level-2 clusters created in the data generating models. While Kim et al. (2021) generated their data with 30 level-2 clusters and Meyers and Beretvas (2006) and Ye and Daniel (2017) generated their data with 30 and 50 clusters, the current simulation study generated comparatively fewer level-2 clusters in certain scenarios. Specifically, conditions generated with 10, 20, and 30 repeated cross-sections resulted in 10, 20, and 30 period-level clusters. Cohort-level cluster sizes depended on the number of repeated cross-sections used to generate the data. Datasets generated with 10 repeated cross-sections resulted in 28 three-year cohort clusters, 17 five-year cohort clusters, and nine ten-year cohort clusters. Datasets generated with 20 repeated cross-sections resulted in 32 three-year cohort clusters, 20 five-year cohort clusters, and ten ten-year cohort clusters. Datasets generated with 30 repeated cross-sections resulted in 39 three-year cohort clusters, 24 five-year cohort clusters, and 12 ten-year cohort clusters. It may therefore be the case that these small cluster sample sizes amplified the impact of the variability in the period and cohort effects to influence the accuracy of the parameters in the study conditions discussed above (Stegmueller, 2013).

The Use of Model Fit Indices to Identify a Cohort Selection Mechanism

This dissertation investigated the performance of the AIC, BIC, AICC, and BICC model fit indices to correctly identify the cohort selection mechanism underlying the structure of the data used in the HAPC-CCREM. The results of the simulation study found that all four indices performed extremely well in identifying the cohort selection mechanisms used in the data generating models. Based on previous simulation research about the performance of model fit indices in cross-classified random effects models (Beretvas & Murphy, 2013), it was anticipated that modifications to the AIC and BIC fit indices may result in better performance, but this study found that these corrections were not necessary. Indeed, there was no practically meaningful difference in the performance of the fit indices across all the study conditions evaluated in the Monte Carlo simulation. The successful performance of these fit indices is likely attributable to the model misspecification issues created by the cohort selection mechanisms discussed above. The most plausible explanation is that the cohort groupings underlying the dataset create such a unique data structure that the fit indices can easily target when the cohort groupings used in the estimating model do not match the cohort groupings underlying the dataset (Reither et al., 2015; Wu et al., 2009; Yang & Land, 2013, pp. 15–19).

Implications of Findings

The findings of this dissertation have several implications for applied researchers seeking to use the HAPC-CCREM in practice. First, the number of survey years used in the analysis should be of concern when seeking to employ the model. The findings of previous research (Yang, 2006) certainly indicate that the model should not be estimated on datasets with five repeated cross-sections under the simulation conditions investigated. If the impact of a period-level predictor is of interest, the findings of this dissertation indicate that 10 repeated cross-sections may be inadequate to accurately estimate the coefficient of the period-level predictor, but 20 repeated cross-sections may be sufficient. If a period-level predictor is not to be included in the model, the findings of this simulation study in concert with previous research (Yang, 2006) indicate that the model parameters can be accurately estimated with 15, 20, or 30 repeated cross-sections under the conditions investigated in the simulations.

Importantly, several published studies have estimated the HAPC-CCREM on datasets with less than 15 repeated cross-sections. For example, gender egalitarianism was examined using three repeated cross-sections (F. Pampel, 2011a); anti-immigrant attitudes were examined using four repeated cross-sections (Gorodzeisky & Semyonov, 2018); self-rated health (Delaruelle et al., 2015), fear of crime among older adults (Koeber & Oberwittler, 2019), and attendance at protests and the signing of petitions (Caren et al., 2011) was examined using 8 repeated cross-sections of data; gender differences in political participation were examined using 11 repeated-cross sections (Andrew S. Fullerton & Stern, 2013); gender differences in attitudes toward racial equality was examined using 12 repeated cross-sections (A. K. Clark, 2017), to name a few (see Appendix A for more studies). Findings of these studies may need to be taken with caution considering the low number of survey years employed in the analyses. It may be advisable to re-estimate the models used in these publications if the questions used in the analyses have continued to be asked since the time of the analysis, where additional years of data were collected by the survey administrators.

The simulation study results related to the effect of the cohort selection mechanism on the accuracy of estimates in the HAPC-CCREM and the use of model fit indices to correctly identify a cohort selection mechanism also have meaningful implications for applied researchers. In this study it was found that when the cohort selection mechanism used in the analysis matched the underlying data structure, the various model parameters were generally estimated accurately. However, when the cohort groupings used did not match, many of the model parameters were inaccurately estimated, especially when 10-year cohort groupings were used in the estimation. Fortunately, this study found that model fit indices commonly employed in most statistical software performed extremely well in identifying the cohort groupings underlying the data. Accordingly, applied researchers could take advantage of these model fit indices early on in the estimation of their model to help ensure that they have not mis-specified the cohort grouping.

Specifically, several competing baseline models with only the age effects at level-1 could be estimated using several different cohort groupings, and the one with the lowest model fit index would indicate the cohort grouping closest to the underlying data structure. In turn, the cohort grouping with the best model fit index would then be used in a conditional model, where additional predictor variables at the individual, period, and cohort levels would be included.

The findings of this dissertation also have important implications for those who fund and manage the datasets used in HAPC-CCREM analyses. To date, the majority of studies using the HAPC-CCREM in practice have utilized the General Social Survey (GSS). The GSS is perhaps the most well-known and frequently used dataset among sociologists, especially given that many items have consistently been asked to respondents using the same question format for more than thirty years. The findings of this dissertation related to the influence of the number of repeated cross-sections on the accuracy of parameters in the HAPC-CCREM should convince GSS administrators to continue to ask the same questions in the same format. If current GSS questions were removed or reworded, not only would historical trend analysis be compromised, but it would be many years before those interested in APC analyses on the new questions using HAPC-CCREM could be employed. For example, the finding of this dissertation that compared to 10 repeated cross-sections, 20 repeated cross-sections produced more accurate coefficient estimates of a period-level predictor would mean that if a new outcome variable is introduced, it would likely be at least 40 years before the HAPC-CCREM could be estimated, considering that a new GSS cross-section is conducted every two calendar years.

Importantly, there are also implications for other datasets as well. For example, comparatively smaller datasets with fewer years of data than the GSS have been utilized in HAPC-CCREM analyses such as the National Travel Survey in England, the European Social Survey, the American National Election Survey, the International Social Survey Program, and the American Public Opinion and Foreign Policy Study (see Appendix A). Administrators and funders of these datasets should be aware that social scientists are using them for HAPC-CCREM analyses and should therefore consider continuing to ask the questions in the same format and for a considerable amount of time to come. Doing so would likely enable accurate model estimates to be obtained in the HAPC-CCREM once the question was consistently asked in 15 repeated cross-sections. A plausible alternative would be for administrators of these datasets to consider releasing a more detailed indicator for when the participant was surveyed, for example the survey month, instead of the survey year. Doing so would readily increase the number of period-level clusters in the analysis and make the employment of HAPC-CCREM analyses more quickly achievable.

Finally, the findings of this dissertation have important implications for life course theory. Decades of prior research has been concerned with solving the identification issue or finding convenient workarounds to the linear dependency between age, period, and cohort effects. When initially introduced, the HAPC-CCREM was mainly put forth as a viable solution to accurately and simultaneously estimate age, period, and cohort effects in one model (Yang & Land, 2006). Most of the applied literature using the model to date has mainly been concerned with accounting for the age, period, and cohort effects in concert with various individual-level predictor variables. However, a growing number of studies are beginning to include predictor variables at the period and cohort levels to examine how they might influence individual-level outcomes. Under the conditions investigated, the findings of this dissertation indicated that for datasets with 20 or more repeated cross-sections, period and cohort level predictors could be accurately estimated, especially when the cohort selection mechanism in the estimating model matched the data generating model.

Applied researchers and life course theorists should consider taking advantage of the multilevel modeling framework offered by the HAPC-CCREM to answer new and more complex questions about age, period, and cohort processes by including relevant predictors at the period and cohort levels. For example, previous HAPC-CCREM analyses have examined individual-level predictors of attitudes towards gender egalitarianism while controlling for differences in periods and cohorts (A. K. Clark, 2017; Donnelly et al., 2016). These analyses could be updated and expanded to include period-level predictors for the percentage of women entering the paid labor force over time or the percentage of women earning college degrees across cohorts. Importantly, measures occurring at the period and cohort level could be linked from other datasets beyond the one used for the individual-level data, a process known as data integration. For example, individual-level measures from the General Social Survey could be linked to period and cohort level information from the World Bank, American Community Survey, or the Current Population Survey. Such novel use of various datasets linked together would greatly expand the types of research questions life course theorists could answer using the HAPC-CCREM, and potentially further enhance our understanding of the nuanced ways that age, period, and cohort effects influence individual-level outcomes and longitudinal social change. At the same time, however, life course scholars answering these types of questions should carefully consider the inclusion of period-level predictors in their model if high variability is occurring for both the period and cohort effects, as this dissertation indicated that in those scenarios the period-level predictor tended to be overestimated under the simulation conditions examined.

Limitations and Areas for Future Research

As with any study, this dissertation has important limitations that should be considered. Research from a variety of disciplines indicate that human capacity is limited in its ability to cognitively process, store, and understand complex information, especially when many factors are interacting with one another (Cowan, 2001; Saaty & Ozdemir, 2003; Sterman, 2006). Therefore, in Monte Carlo simulation studies boundaries must be drawn on the number of study conditions under investigation and the levels of these study conditions, especially when fully crossed study designs are employed, as in this dissertation (Beaujean, 2018). One of the main areas of investigation in this dissertation was the impact that the number of survey years has on the accuracy of estimates in the HAPC-CCREM, specifically focusing on 10, 20, and 30 years of data availability. As discussed in Chapter 2, these specific values were selected for analysis mainly to determine if a dataset with the number of survey years between five and 15 could accurately estimate the model parameters, and to further examine how the model performed with larger datasets, especially in the presence of a period predictor.

One of the interesting findings of this dissertation was that datasets with 10 repeated cross-sections exhibited varying degrees of bias in model estimates, and that if the impact of a period predictor was of interest, at least 20 repeated cross-sections of data would be recommended to accurately estimate the coefficient of the period-level parameter. While these findings are helpful, future research could more fully explore the required minimum number of survey years to accurately estimate the parameters in the HAPC-CCREM. For the reasons discussed above, the number of survey years in this dissertation was limited to 10, 20, and 30 repeated cross-sections. It would be useful, however, to conduct a simulation study that expanded on these levels to determine a more specific number of survey years to use to estimate the model without a period predictor (for example, 10, 11, 12, 13, 14, or 15 years), as well as a more specific number of survey years recommended to estimate a period predictor (for example, somewhere between 10 to 20 years of data).

This study was also limited considering several other factors that could have been explored but were left uninvestigated to manage the total number of manipulated conditions. For example, the within-cluster sample sizes were not investigated for the number of repeated cross-sections available in the analysis. The simulation study used in this dissertation was designed to reflect the overall sample size of the General Social Survey, which is roughly 1500 respondents in each repeated cross section. However, as mentioned in Chapter 3 the GSS reduces survey administration costs by employing a split ballot design so that not all respondents are asked all questions. The split ballot design, in addition to participants refusing to answer a particular question or responding “don’t know” to a particular question very quickly reduces the number of participants in each repeated cross-section when list-wise deletion procedures are utilized to derive the final sample for the analysis. Future Monte Carlo simulation studies on the HAPC-CCREM should therefore consider the impact that smaller within-cluster sample sizes may have on the accuracy of the model’s estimates. This would be especially interesting for using a 3-year cohort selection mechanism, which would result in a larger number of cohort clusters but a smaller sample size within the cohort clusters.

While one of the strengths of this study was that it evaluated the inclusion of a period-level predictor in the model, which was unexamined by previous research, it did not consider many other modeling choices that could be made by applied researchers. As discussed in the previous section, the multilevel modeling framework of the HAPC-CCREM will allow life course researchers to examine more complex and nuanced period- and cohort-level processes that impact social change on outcomes of interest. In particular, researchers could specify individual-level predictors that randomly vary across periods and/or cohorts and could also specify cross-level interaction effects between individual-level variables and the period- and cohort-level predictors. Future simulation studies of the HAPC-CCREM could build on the current research by examining the ability of the model to accurately estimate these types of effects.

Future simulation studies could also evaluate other cohort selection mechanisms that may be of interest to applied researchers. This dissertation focused on the most commonly used cohort groupings (3-year, 5-year, and 10-year) in previously published HAPC-CCREM analyses. However, at least two studies in the literature utilized generational theory to form their cohort groupings (see Kowske et al., 2010 and Shu & Meagher, 2018). Generational theory is a subfield of life course theory that specifically examines how very large generations of individuals influence outcomes of interest (Mannheim, 1952; Okros, 2020), such as the GI Generation (those born between 1901-1924), Baby Boomers (those born between 1943-1960), Gen Xers (those born between 1961-1981), and Millenials (those born between 1982-2003). Using these cohort groupings in the model may have important impacts on the accuracy of the model estimates. Speculatively, the findings of this dissertation would imply that if the underlying structure of the data used in this model was formed by these groupings, the model would produce accurate effects as long as the same grouping was specified in the model itself (i.e., the cohort selection mechanism in the data generating model matched the estimating model). However, if this was not the case, using these cohorts would likely cause biased model estimates, especially given that they would result in a very small number of cohort clusters used by the model since each generation encompasses a very large number of birth cohorts. Nonetheless, future simulation studies could explicitly test this assumption.

Additionally, it will remain important to revisit and test other specifications for cohort groupings as our understanding of social processes is updated or advanced, especially if it influences model selection and the variability exhibited across cohorts. Speculatively, for example, the continued rapid advancement of technology may produce cohort effects that require cohort groupings used in age-period-cohort analyses that differ from current practice. In this example, it may be the case that rapid technological developments result in cohort effects that are more variable among newer (that is, younger) individuals compared to more stable effects in older groups. Such a phenomenon could be modeled in HAPC-CCREM analyses by blending a modeling approach of narrower cohort widths for the newer cohorts (perhaps 3-year or 5-year birth cohorts) and using wider cohort widths informed by generational theory for the older birth cohorts. If these types of model specification procedures are of interest to applied researchers, they should be evaluated as a cohort selection mechanism in future simulation studies.

Another limitation of this dissertation is that the findings of the simulation study only apply to HAPC-CCREMs estimated with continuous outcomes. Monte Carlo simulations conducted on multilevel models in general consistently indicate that compared to continuous outcomes, models with non-continuous outcomes require greater sample sizes to accurately estimate model parameters, especially for fixed effects at level-2 and variance components at level-1 and level-2 (McNeish & Harring, 2017; Moineddin et al., 2007; Schoeneberger, 2016). While many of the studies using the HAPC-CCREM in the applied literature to date have used continuous outcomes, a growing number of studies have examined binary outcomes (for example, see Attell, 2020 and Keyes et al., 2011) and ordered-categorical outcomes (for example, see Yang, 2008 and Zhang, 2017). It would therefore be worthy of investigation to examine the ability of the HAPC-CCREM to accurately estimate parameters with non-continuous dependent variables and if model fit indices for these types of outcomes perform as well as they did in the current study in detecting the cohort selection mechanism underlying the data.

Future research on the performance of the HAPC-CCREM could also investigate coverage rates of the 95% confidence intervals for the fixed effects using different methods to calculate the confidence intervals. In this study, confidence intervals were calculated using the Wald method, as discussed in Chapter 3. The Wald method is commonly utilized to calculate confidence intervals given its computational efficiency. However, it relies on the asymptotic normality properties of the maximum likelihood estimate of the coefficient and its standard error, meaning that in small sample sizes the Wald method may produce overly narrow confidence intervals (Royston, 2007). Therefore, some studies have proposed other methods to calculate confidence intervals that are robust to this issue, such as profile likelihood estimation and bootstrapping, that would result in wider confidence intervals (Carpenter & Bithell, 2000; Longford, 2000; Venzon & Moolgavkar, 1988). It would therefore be worthy of future research to examine if the low coverage rates found for some of the conditions in this study could be improved using a different calculation of the 95% confidence interval.

One final area for future methodological work related to the HAPC-CCREM would be to establish a comprehensive framework for building and reporting the model. Such frameworks exist for more traditional two-level models (Bliese & Ployhart, 2002; Grilli & Rampichini, 2018; McCoach, 2019; McCoach & Kaniskan, 2010; Niehaus et al., 2014), but currently do not exist for the HAPC-CCREM. Findings from this dissertation related to the use of model fit indices to identify an ideal cohort selection mechanism could be combined with previously existing methodological research related to centering variables in the model (Yang & Land, 2006) and examining the significance of the period and cohort effects (Frenk et al., 2013). A comprehensive model building framework would consider how these model building choices should work together and in what steps they should be performed to aid applied researchers in carefully and thoughtfully executing their models. Such a framework would also have the potential to help peer reviewers of research studies employing the HAPC-CCREM evaluate if assumptions of the model had been met and if critical elements of the model were appropriately reported in manuscripts submitted to journals.

REFERENCES

Akaike, H. (1998). Information Theory and an Extension of the Maximum Likelihood Principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected Papers of Hirotugu Akaike* (pp. 199–213). Springer New York. https://doi.org/10.1007/978-1-4612-1694-0_15

Allison, P. D. (2009). *Fixed effects regression models*. Sage.

Alwin, D. F., & McCammon, R. J. (2003). Generations, cohorts, and social change. In *Handbook of the life course* (pp. 23–49). Springer.

An, Z., Heinen, E., & Watling, D. (2021). When you are born matters: An age-period-cohort analysis of multimodality. *Travel Behaviour and Society*, *22*, 129–145. https://doi.org/10.1016/j.tbs.2020.09.002

Anderson, A. L., Lytle, R., & Schwadel, P. (2017). Age, period, and cohort effects on death penalty attitudes in the United States, 1974–2014. *Criminology*, *55*(4), 833–868. https://doi.org/10.1111/1745-9125.12160

Attell, B. K. (2020). Changing Attitudes Toward Euthanasia and Suicide for Terminally Ill Persons, 1977 to 2016: An Age-Period-Cohort Analysis. *OMEGA - Journal of Death and Dying*, *80*(3), 355–379. https://doi.org/10.1177/0030222817729612

Bardo, A. R. (2017). A life course model for a domains-of-life approach to happiness: Evidence from the United States. *Advances in Life Course Research*, *33*, 11–22. https://doi.org/10.1016/j.alcr.2017.06.002

Bardo, A. R., Lynch, S. M., & Land, K. C. (2017). The importance of the Baby Boom cohort and the Great Recession in understanding age, period, and cohort patterns in happiness. *Social Psychological and Personality Science*, *8*(3), 341–350. https://doi.org/10.1177/1948550616673874

Baunach, D. M. (2012). Changing Same-Sex Marriage Attitudes in America from 1988 Through 2010. *Public Opinion Quarterly*, *76*(2), 364–378. https://doi.org/10.1093/poq/nfs022

Beaujean, A. A. (2018). Simulating data for clinical research: A tutorial. *Journal of Psychoeducational Assessment*, *36*(1), 7–20. https://doi.org/10.1177/0734282917690302

Bell, A., & Jones, K. (2013). The impossibility of separating age, period and cohort effects. *Social Science & Medicine*, *93*, 163–165. https://doi.org/10.1016/j.socscimed.2013.04.029

Bell, A., & Jones, K. (2014a). Another 'futile quest'? A simulation study of Yang and Land's hierarchical age-period-cohort model. *Demographic Research*, *30*, 333–360. JSTOR.

Bell, A., & Jones, K. (2014b). Don't birth cohorts matter? A commentary and simulation exercise on Reither, Hauser, and Yang's (2009) age–period–cohort study of obesity. *Social Science & Medicine*, *101*, 176–180. https://doi.org/10.1016/j.socscimed.2013.09.004

Bell, A., & Jones, K. (2014c). Current practice in the modelling of age, period and cohort effects with panel data: A commentary on Tawfik et al. (2012), Clarke et al. (2009), and McCulloch (2012). *Quality & Quantity*, *48*(4), 2089–2095. https://doi.org/10.1007/s11135-013-9881-x

Bell, A., & Jones, K. (2015). Should age-period-cohort analysts accept innovation without scrutiny? A response to Reither, Masters, Yang, Powers, Zheng and Land. *Social Science & Medicine*, *128*, 331–333. https://doi.org/10.1016/j.socscimed.2015.01.040

Beretvas, N. S. (2008). Cross-classified random effects models. In A. A. O'Connell & B. D. McCoach (Eds.), *Multilevel modeling of educational data* (pp. 140–180). Information Age Publishing.

Beretvas, N. S. (2011). Cross-classified and multiple-membership models. In J. J. Hox & J. K. Roberts (Eds.), *Handbook of advanced multilevel analysis* (pp. 313–334). Routledge.

Beretvas, N. S., & Murphy, D., L. (2013). An evaluation of information criteria use for correct cross-classified random effects model selection. *The Journal of Experimental Education*, *81*(4), 429–463. https://doi.org/10.1080/00220973.2012.745467

Bernardi, L., Huinink, J., & Settersten, R. (2020). Linking demographic change and the lifecourse: Insights from the life course cube. In *Handbook on demographic change and the lifecourse* (pp. 7–21). Edward Elgar Publishing. https://doi.org/10.4337/9781788974875.00008

Bernardi, L., Huinink, J., & Settersten, R. A. (2019). The life course cube: A tool for studying lives. *Theoretical and Methodological Frontiers in Life Course Research*, *41*, 100258. https://doi.org/10.1016/j.alcr.2018.11.004

Blalock, H. M. (1967). Status inconsistency, social mobility, status integration and structural effects. *American Sociological Review*, *32*(5), 790–801. JSTOR. https://doi.org/10.2307/2092026

Bliese, P. D., & Ployhart, R. E. (2002). Growth modeling using random coefficient models: Model building, testing, and illustrations. *Organizational Research Methods*, *5*(4), 362–387. https://doi.org/10.1177/109442802237116

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*(2), 144–152. https://doi.org/10.1111/j.2044-8317.1978.tb00581.x

Brady, H. E., & Johnston, R. (2015). Repeated cross-sections in survey data. In *Emerging Trends in the Social and Behavioral Sciences* (pp. 1–18). American Cancer Society. https://doi.org/10.1002/9781118900772.etrds0284

Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine*, *25*(24), 4279–4292. https://doi.org/10.1002/sim.2673

Caren, N., Ghoshal, R. A., & Ribas, V. (2011). A social movement generation: Cohort and period trends in protest attendance and petition signing. *American Sociological Review*, *76*(1), 125–151. https://doi.org/10.1177/0003122410395369

Carlisle, J. E., & Clark, A. K. (2018). Green for God: Religion and environmentalism by cohort and time. *Environment and Behavior*, *50*(2), 213–241. https://doi.org/10.1177/0013916517693356

Carpenter, J., & Bithell, J. (2000). Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. *Statistics in Medicine*, *19*(9), 1141–1164. https://doi.org/10.1002/(SICI)1097-0258(20000515)19:9<1141::AID-SIM479>3.0.CO;2-F

Carson, G. D. (2013). Fundamentals of hierarchical linear and multilevel modelling. In G. D. Carson (Ed.), *Hierarchical linear modeling: Guide and applications* (pp. 3–25). Sage.

Clark, A., Justwan, F., Carlisle, J. E., & Clark, M. (2019). Polarization politics and hopes for a green agenda in the United States. *Environmental Politics*, 1–27. https://doi.org/10.1080/09644016.2019.1654238

Clark, A. K. (2017). Updating the Gender Gap(s): A Multilevel Approach to What Underpins Changing Cultural Attitudes. *Politics & Gender*, *13*(1), 26–56. Cambridge Core. https://doi.org/10.1017/S1743923X16000520

Costanza, D. P., Darrow, J. B., Yost, A. B., & Severt, J. B. (2017). A review of analytical methods used to study generational differences: Strengths and limitations. *Work, Aging and Retirement*, *3*(2), 149–165. https://doi.org/10.1093/workar/wax002

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. Cambridge Core. https://doi.org/10.1017/S0140525X01003922

Delaruelle, K., Buffel, V., & Bracke, P. (2015). Educational expansion and the education gradient in health: A hierarchical age-period-cohort analysis. *Social Science & Medicine*, *145*, 79–88. https://doi.org/10.1016/j.socscimed.2015.09.040

Dennis, B., Ponciano, J. M., Taper, M. L., & Lele, S. R. (2019). Errors in Statistical Inference Under Model Misspecification: Evidence, Hypothesis Testing, and AIC. *Frontiers in Ecology and Evolution*, *7*, 372. https://doi.org/10.3389/fevo.2019.00372

Donnelly, K., Twenge, J. M., Clark, M. A., Shaikh, S. K., Beiler-May, A., & Carter, N. T. (2016). Attitudes toward women's work and family roles in the United States, 1976–2013. *Psychology of Women Quarterly*, *40*(1), 41–54. https://doi.org/10.1177/0361684315590774

Ebbes, P., Böckenholt, U., & Wedel, M. (2004). Regressor and random-effects dependencies in multilevel models. *Statistica Neerlandica*, *58*(2), 161–178. https://doi.org/10.1046/j.0039-0402.2003.00254.x

Eisenstein, M. A., Clark, A. K., & Jelen, T. G. (2017). Political tolerance and religion: An age-period-cohort analysis, 1984–2014. *Review of Religious Research*, *59*(3), 395–418. https://doi.org/10.1007/s13644-017-0295-4

Elder, G. (1994). Time, Human Agency, and Social Change: Perspectives on the Life Course. *Social Psychology Quarterly*, *57*(1), 4–15.

Elder, G. (1998). The Life Course as Developmental Theory. *Child Development*, *69*(1), 1–12.

Elder, G., Johnson, M., & Crosnoe, R. (2003). The Emergence and Development of Life Course Theory. In *Handbook of the Life Course* (pp. 3–19). Kluswer Academic Publishers.

Elias, V. L., Fullerton, A. S., & Simpson, J. M. (2015). Long-term changes in attitudes toward premarital sex in the United States: Reexamining the role of cohort replacement. *The Journal of Sex Research*, *52*(2), 129–139. https://doi.org/10.1080/00224499.2013.798610

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*(2), 121–138.

Fairbrother, M. (2014). Two Multilevel Modeling Techniques for Analyzing Comparative Longitudinal Survey Datasets. *Political Science Research and Methods*, *2*(1), 119–140. https://doi.org/10.1017/psrm.2013.24

Fielding, A. (2004). The role of the Hausman test and whether higher level effects should be treated as random or fixed. *Multilevel Modelling Newsletter*, *16*(2), 3–9.

Fielding, A., & Goldstein, H. (2006). *Cross-classified and multiple membership structures in multilevel models:* (p. 69).

Fienberg, S. E., & Mason, W. M. (1979). Identification and estimation of age-period-cohort models in the analysis of discrete archival data. *Sociological Methodology*, *10*, 1–67. JSTOR. https://doi.org/10.2307/270764

Firebaugh, G. (1997). *Analyzing repeated surveys*. Sage.

Fosse, E., & Winship, C. (2019). Analyzing age-period-cohort data: A review and critique. *Annual Review of Sociology*, *45*(1), annurev-soc-073018-022616. https://doi.org/10.1146/annurev-soc-073018-022616

Frenk, S. M., Yang, Y. C., & Land, K. C. (2013). Assessing the significance of cohort and period effects in hierarchical age-period-cohort models: Applications to verbal test scores and voter turnout in U.S. presidential elections. *Social Forces*, *92*(1), 221–248. https://doi.org/10.1093/sf/sot066

Fullerton, A. S., & Stern, M. J. (2010). Explaining the persistence and eventual decline of the gender gap in voter registration and turnout in the American South, 1956-1980. *Social Science History*, *34*(2), 129–169. https://doi.org/10.1215/01455532-2009-023

Fullerton, Andrew S., & Dixon, J. C. (2010a). Generational conflict or methodological artifact? Reconsidering the relationship between age and policy attitudes in the U.S., 1984–2008. *Public Opinion Quarterly*, *74*(4), 643–673. https://doi.org/10.1093/poq/nfq043

Fullerton, Andrew S., & Dixon, J. C. (2010b). Generational Conflict Or Methodological Artifact?: Reconsidering the Relationship between Age and Policy Attitudes in the U.S., 1984–2008. *Public Opinion Quarterly*, *74*(4), 643–673. https://doi.org/10.1093/poq/nfq043

Fullerton, Andrew S., & Stern, M. J. (2013). Racial differences in the gender gap in political participation in the American South, 1952–2004. *Social Science History*, *37*(2), 145–176. https://doi.org/10.1017/S0145553200010622

Gauchat, G. (2012). Politicization of science in the public sphere: A study of public trust in the United States, 1974 to 2010. *American Sociological Review*, *77*(2), 167–187. https://doi.org/10.1177/0003122412438225

George, L. K. (1993). Sociological Perspectives on Life Transitions. *Annual Review of Sociology*, *19*, 353–373.

Glenn, N. D. (1976). Cohort analysts' futile quest: Statistical attempts to separate age, period and cohort effects. *American Sociological Review*, *41*(5), 900–904. JSTOR. https://doi.org/10.2307/2094738

Glenn, N. D. (2002). Distinguishing age, period, and cohort effects. In *Handbook of the Life Course* (pp. 465–476). Kluswer Academic Publishers.

Glenn, N. D. (2005). *Cohort analysis*. Sage.

Goldstein, H. (1994). Multilevel cross-classified models. *Sociological Methods & Research*, *22*(3), 364–375. https://doi.org/10.1177/0049124194022003005

Gorodzeisky, A., & Semyonov, M. (2018). Competitive threat and temporal change in anti-immigrant sentiment: Insights from a hierarchical age-period-cohort model. *Social Science Research*, *73*, 31–44. https://doi.org/10.1016/j.ssresearch.2018.03.013

Grilli, L., & Rampichini, C. (2015). Specification of random effects in multilevel models: A review. *Quality & Quantity*, *49*(3), 967–976. https://doi.org/10.1007/s11135-014-0060-5

Grilli, L., & Rampichini, C. (2018). A handful of critical choices in multilevel modelling. *Boletin de Estadistica e Investigacion Operativa*, *34*(1), 7–24.

Gurka, M. J. (2006). Selecting the Best Linear Mixed Model Under REML. *The American Statistician*, *60*(1), 19–26. https://doi.org/10.1198/000313006X90396

Harding, D. J. (2009). Recent advances in age-period-cohort analysis. A commentary on Dregan and Armstrong, and on Reither, Hauser and Yang. *Social Science & Medicine*, *69*(10), 1449–1451. https://doi.org/10.1016/j.socscimed.2009.08.034

Heckman, J., & Robb, R. (1985). Using longitudinal data to estimate age, period and cohort effects in earnings equations. In William M. Mason & S. E. Fienberg (Eds.), *Cohort Analysis in Social Research: Beyond the Identification Problem* (pp. 137–150). Springer New York. https://doi.org/10.1007/978-1-4613-8536-3_5

Heinz, W., & Kruger, H. (2001). Life Course: Innovations and Challenges for Social Research. *Current Sociology*, *49*(2), 29–45.

Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management*, *24*(5), 623–641. https://doi.org/10.1177/014920639802400504

Hoogland, J. J., & Boomsma, A. (1998). Robustness Studies in Covariance Structure Modeling: An Overview and a Meta-Analysis. *Sociological Methods & Research*, *26*(3), 329–367. https://doi.org/10.1177/0049124198026003003

Horowitz, J. (2015). Doing less with more: Cohorts, education, and civic participation in America. *Social Forces*, *94*(2), 747–774. https://doi.org/10.1093/sf/sov065

Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*(2), 297–307. https://doi.org/10.1093/biomet/76.2.297

Johnson, A. L. (2021). Changes in Mental Health and Treatment, 1997–2017. *Journal of Health and Social Behavior*, 0022146520984136. https://doi.org/10.1177/0022146520984136

Johnson, E. W., & Schwadel, P. (2019). Political Polarization and Long-Term Change in Public Support for Environmental Spending. *Social Forces*, *98*(2), 915–941. https://doi.org/10.1093/sf/soy124

Keyes, K. M., Gary, D., O'Malley, P. M., Hamilton, A., & Schulenberg, J. (2019). Recent increases in depressive symptoms among US adolescents: Trends from 1991 to 2018.

Social Psychiatry and Psychiatric Epidemiology, *54*(8), 987–996. https://doi.org/10.1007/s00127-019-01697-8

Keyes, K. M., & Li, G. (2012). Age–period–cohort modeling. In G. Li & S. P. Baker (Eds.), *Injury Research* (pp. 409–426). Springer US. https://doi.org/10.1007/978-1-4614-1599-2_22

Keyes, K. M., Schulenberg, J. E., O'Malley, P. M., Johnston, L. D., Bachman, J. G., Li, G., & Hasin, D. (2011). The social norms of birth cohorts and adolescent marijuana use in the United States, 1976–2007. *Addiction*, *106*(10), 1790–1800. https://doi.org/10.1111/j.1360-0443.2011.03485.x

Keyes, K. M., Schulenberg, J. E., O'Malley, P. M., Johnston, L. D., Bachman, J. G., Li, G., & Hasin, D. (2012). Birth cohort effects on adolescent alcohol use: The influence of social norms from 1976 to 2007. *Archives of General Psychiatry*, *69*(12), 1304–1313. https://doi.org/10.1001/archgenpsychiatry.2012.787

Kim, S., Jeong, Y., & Hong, S. (2021). The Impact of Ignoring a Crossed Factor in Cross-Classified Multilevel Modeling. *Frontiers in Psychology*, *12*, 567. https://doi.org/10.3389/fpsyg.2021.637645

Koeber, G., & Oberwittler, D. (2019). How older people became less afraid of crime—An age-period-cohort analysis using repeated cross-sectional survey data. *Social Science Research*, *79*, 211–225. https://doi.org/10.1016/j.ssresearch.2018.10.010

Kowske, B. J., Rasch, R., & Wiley, J. (2010). Millennials' (lack of) attitude problem: An empirical examination of generational effects on work attitudes. *Journal of Business and Psychology*, *25*(2), 265–279. https://doi.org/10.1007/s10869-010-9171-8

Kraemer, H. C., Yesavage, J. A., Taylor, J. L., & Kupfer, D. (2000). How Can We Learn About Developmental Processes From Cross-Sectional Studies, or Can We? *American Journal of Psychiatry*, *157*(2), 163–171. https://doi.org/10.1176/appi.ajp.157.2.163

Landes, S. D., & Settersten, R. A. (2019). The inseparability of human agency and linked lives. *Advances in Life Course Research*, *42*, 100306. https://doi.org/10.1016/j.alcr.2019.100306

Lee, Y. R., & Hong, S. (2019). The impact of omitting random interaction effects in cross-classified random effect modeling. *The Journal of Experimental Education*, *87*(4), 641–660. https://doi.org/10.1080/00220973.2018.1507985

Longford, N. T. (2000). On Estimating Standard Errors in Multilevel Analysis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *49*(3), 389–398. https://doi.org/10.1111/1467-9884.00242

Luo, W., & Kwok, O. (2009). The impacts of ignoring a crossed factor in analyzing cross-classified data. *Multivariate Behavioral Research*, *44*(2), 182–212. https://doi.org/10.1080/00273170902794214

Lutz, W. (2012). Demographic Metabolism: A Predictive Theory of Socioeconomic Change. *Population and Development Review*, *38*, 283–301.

Maas, C., & Hox, J. (2002). *Robustness of multilevel parameter estimates against small sample sizes* [Unpublished paper]. http://www.joophox.net/papers/p090101.pdf

Mannheim, K. (1952). The problem of generations. In *Essays on the sociology of knowledge* (pp. 163–195). Routledge.

Mason, K. O., Mason, W. M., Winsborough, H. H., & Poole, W. K. (1973). Some methodological issues in cohort analysis of archival data. *American Sociological Review*, *38*(2), 242–258. JSTOR. https://doi.org/10.2307/2094398

Mason, W., & Fienberg, S. (1985). *Cohort Analysis in Social Research: Beyond the Identification Problem*. Springer.

Mason, W. M., & Fienberg, S. (Eds.). (1985). *Cohort analysis in social research: Beyond the identification problem*. Springer-Verlag. https://www.springer.com/gp/book/9781461385387

Masters, R. K. (2012). Uncrossing the U.S. black-white mortality crossover: The role of cohort forces in life course mortality risk. *Demography*, *49*(3), 773–796. https://doi.org/10.1007/s13524-012-0107-y

Masters, R. K., Hummer, R. A., & Powers, D. A. (2012). Educational differences in U.S. adult mortality: A cohort perspective. *American Sociological Review*, *77*(4), 548–572. https://doi.org/10.1177/0003122412451019

Mayer, K. U. (2009). New Directions in Life Course Research. *Annual Review of Sociology*, *35*(1), 413–433. https://doi.org/10.1146/annurev.soc.34.040507.134619

McCoach, B. (2019). Multilevel modeling. In *The reviewer's guide to quantitative methods in the social sciences* (2nd Edition, pp. 292–312). Routledge.

McCoach, B., & Kaniskan, B. (2010). Using time-varying covariates in multilevel growth models. *Frontiers in Psychology*, *1*. https://doi.org/10.3389/fpsyg.2010.00017

McNeish, D. (2014). Modeling sparsely clustered data: Design-based, model-based, and single-level methods. *Psychological Methods*, *19*(4), 552–563. APA PsycArticles. https://doi.org/10.1037/met0000024

McNeish, D., & Harring, J. (2017). Clustered data with small sample sizes: Comparing the performance of model-based and design-based approaches. *Communications in Statistics - Simulation and Computation*, *46*(2), 855–869. https://doi.org/10.1080/03610918.2014.983648

McNeish, D., & Stapleton, L. (2016a). The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychology Review*, *28*(2), 295–314. https://doi.org/10.1007/s10648-014-9287-x

McNeish, D., & Stapleton, L. (2016b). Modeling Clustered Data with Very Few Clusters. *Multivariate Behavioral Research*, *51*(4), 495–518. https://doi.org/10.1080/00273171.2016.1167008

Meuleman, B. (2019). The economic context of solidarity. Period vs. Cohort differences in support for income redistribution in Britain and the United States. *European Societies*, *21*(5), 774–801. https://doi.org/10.1080/14616696.2019.1616792

Meyers, J. L., & Beretvas, S. N. (2006). The impact of inappropriate modeling of cross-classified data structures. *Multivariate Behavioral Research*, *41*(4), 473–497. https://doi.org/10.1207/s15327906mbr4104_3

Moineddin, R., Matheson, F. I., & Glazier, R. H. (2007). A simulation study of sample size for multilevel logistic regression models. *BMC Medical Research Methodology*, *7*(1), 34. https://doi.org/10.1186/1471-2288-7-34

Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, *38*(11), 2074–2102. https://doi.org/10.1002/sim.8086

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*(1), 90–100. https://doi.org/10.1016/S0022-2496(02)00028-7

Niehaus, E., Campbell, C. M., & Inkelas, K. K. (2014). HLM behind the curtain: Unveiling decisions behind the use and interpretation of HLM in higher education research. *Research in Higher Education*, *55*(1), 101–122. https://doi.org/10.1007/s11162-013-9306-7

O'Brien, R. M. (1989). Relative cohort size and age-specific crime rates: An age-period-relative-cohort-size model. *Criminology*, *27*(1), 57–78. https://doi.org/10.1111/j.1745-9125.1989.tb00863.x

O'Brien, R. M. (2000). Age period cohort characteristic models. *Social Science Research*, *29*(1), 123–139. https://doi.org/10.1006/ssre.1999.0656

O'Brien, R. M., Stockard, J., & Isaacson, L. (1999). The enduring effects of cohort characteristics on age-specific homicide rates, 1960-1995. *American Journal of Sociology*, *104*(4), 1061–1095. https://doi.org/10.1086/210136

Okros, A. (2020). Generational Theory and Cohort Analysis. In A. Okros (Ed.), *Harnessing the Potential of Digital Post-Millennials in the Future Workplace* (pp. 33–51). Springer International Publishing. https://doi.org/10.1007/978-3-030-25726-2_2

Paccagnella, O. (2006). Centering or not centering in multilevel models? The role of the group mean and the assessment of group effects. *Evaluation Review*, *30*(1), 66–85. https://doi.org/10.1177/0193841X05275649

Pampel, F. (2011a). Cohort change, diffusion, and support for gender egalitarianism in cross-national perspective. *Demographic Research*, *25*, 667–694. JSTOR.

Pampel, F. (2011b). Cohort changes in the socio-demographic determinants of gender egalitarianism. *Social Forces*, *89*(3), 961–982. https://doi.org/10.1353/sof.2011.0011

Pampel, F. C. (2016). Cohort changes in the social distribution of tolerant sexual attitudes. *Social Forces*, *95*(2), 753–777. https://doi.org/10.1093/sf/sow069

Pampel, F. C., & Hunter, L. M. (2012). Cohort change, diffusion, and support for environmental spending in the United States. *American Journal of Sociology*, *118*(2), 420–448. https://doi.org/10.1086/666506

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2011). Sources of Method Bias in Social Science Research and Recommendations on How to Control It. *Annual Review of Psychology*, *63*(1), 539–569. https://doi.org/10.1146/annurev-psych-120710-100452

Rasbash, J., & Goldstein, H. (1994). Efficient analysis of mixed hierarchical and cross-classified random structures using a multilevel model. *Journal of Educational and Behavioral Statistics*, *19*(4), 337–350. JSTOR. https://doi.org/10.2307/1165397

Raudenbush, S. W. (1993). A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. *Journal of Educational Statistics*, *18*(4), 321–349. JSTOR. https://doi.org/10.2307/1165158

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Sage.

Reither, E. N., Hauser, R. M., & Yang, Y. (2009). Do birth cohorts matter? Age-period-cohort analyses of the obesity epidemic in the United States. *Social Science & Medicine*, *69*(10), 1439–1448. https://doi.org/10.1016/j.socscimed.2009.08.040

Reither, E. N., Masters, R. K., Yang, Y. C., Powers, D. A., Zheng, H., & Land, K. C. (2015). Should age-period-cohort studies return to the methodologies of the 1970s? *Social Science & Medicine*, *128*, 356–365. https://doi.org/10.1016/j.socscimed.2015.01.011

Rosenberger, J. S., & Steiner, M. (2019). Attitudes toward Unauthorized Immigration in the Post-9/11 Era. *Sociological Focus*, *52*(4), 292–308. https://doi.org/10.1080/00380237.2019.1668318

Royston, P. (2007). Profile Likelihood for Estimation and Confidence Intervals. *The Stata Journal*, *7*(3), 376–387. https://doi.org/10.1177/1536867X0700700305

Ruspini, E. (1999). Longitudinal Research and the Analysis of Social Change. *Quality and Quantity*, *33*(3), 219–227. https://doi.org/10.1023/A:1004692619235

Ryder, N. B. (1965). The Cohort as a Concept in the Study of Social Change. *American Sociological Review*, *30*(6), 843–861. JSTOR. https://doi.org/10.2307/2090964

Saaty, T. L., & Ozdemir, M. S. (2003). Why the magic number seven plus or minus two. *Mathematical and Computer Modelling*, *38*(3), 233–244. https://doi.org/10.1016/S0895-7177(03)90083-5

Sanderson, M. R., Semyonov, M., & Gorodzeisky, A. (2021). Declining and splitting: Opposition to immigration in the United States, 1996–2018. *International Journal of Intercultural Relations*, *80*, 27–39. https://doi.org/10.1016/j.ijintrel.2020.11.001

Satterfield-Nash, A., Kotzky, K., Allen, J., Bertolli, J., Moore, C. A., Pereira, I. O., Pessoa, A., Melo, F., Santelli, A. C. F. E. S., Boyle, C. A., & Peacock, G. (2017). Health and Development at Age 19-24 Months of 19 Children Who Were Born with Microcephaly and Laboratory Evidence of Congenital Zika Virus Infection During the 2015 Zika Virus Outbreak—Brazil, 2017. *MMWR. Morbidity and Mortality Weekly Report*, *66*(49), 1347–1351. PubMed. https://doi.org/10.15585/mmwr.mm6649a2

Schoeneberger, J. A. (2016). The Impact of Sample Size and Other Factors When Estimating Multilevel Logistic Models. *The Journal of Experimental Education*, *84*(2), 373–397. https://doi.org/10.1080/00220973.2015.1027805

Schwadel, P. (2010a). Period and cohort effects on religious nonaffiliation and religious disaffiliation: A research note. *Journal for the Scientific Study of Religion*, *49*(2), 311–319. https://doi.org/10.1111/j.1468-5906.2010.01511.x

Schwadel, P. (2010b). Age, period, and cohort effects on U.S. religious service attendance: The declining impact of sex, southern residence, and Catholic affiliation. *Sociology of Religion*, *71*(1), 2–24. https://doi.org/10.1093/socrel/srq005

Schwadel, P. (2013a). Changes in Americans' views of prayer and reading the Bible in public schools: Time periods, birth cohorts, and religious traditions. *Sociological Forum*, *28*(2), 261–282. https://doi.org/10.1111/socf.12018

Schwadel, P. (2013b). Changes in Americans' strength of religious affiliation, 1974–2010. *Sociology of Religion*, *74*(1), 107–128. https://doi.org/10.1093/socrel/srs043

Schwadel, P., & Ellison, C. G. (2017). Period and cohort changes in Americans' support for marijuana legalization: Convergence and divergence across social groups. *The Sociological Quarterly*, *58*(3), 405–428. https://doi.org/10.1080/00380253.2017.1331715

Schwadel, P., & Garneau, C. R. H. (2014). An age–period–cohort analysis of political tolerance in the United States. *The Sociological Quarterly*, *55*(2), 421–452. https://doi.org/10.1111/tsq.12058

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Settersten, R. A., & Mayer, K. U. (1997). The Measurement of Age, Age Structuring, and the Life Course. *Annual Review of Sociology*, *23*, 233–261.

Shi, L., Lu, Y., & Pickett, J. T. (2020). The public salience of crime, 1960–2014: Age–period–cohort and time–series analyses. *Criminology*, *58*(3), 568–593. https://doi.org/10.1111/1745-9125.12248

Shi, Y., Leite, W., & Algina, J. (2010). The impact of omitting the interaction between crossed factors in cross-classified random effects modelling. *British Journal of Mathematical and Statistical Psychology*, *63*(1), 1–15. https://doi.org/10.1348/000711008X398968

Shu, X., & Meagher, K. D. (2018). Beyond the stalled gender revolution: Historical and cohort dynamics in gender attitudes from 1977 to 2016. *Social Forces*, *96*(3), 1243–1274. https://doi.org/10.1093/sf/sox090

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

Smets, K., & Neundorf, A. (2014). The hierarchies of age-period-cohort research: Political context and the development of generational turnout patterns. *Electoral Studies*, *33*, 41–51. https://doi.org/10.1016/j.electstud.2013.06.009

Smith, H., Mason, W., & Fienberg, S. (1982). Estimable functions of age, period, and cohort effects: More chimeras of the age-period-cohort accounting framework: comment on rodgers. *American Sociological Review*, *47*(6), 787–793. JSTOR. https://doi.org/10.2307/2095214

Smith, T., Houst, M., & Marsden, P. (2017). *General Social Survey, 1972-2016 Cumulative File* [Data set]. https://www.icpsr.umich.edu/web/ICPSR/studies/36797

Snijders, T. A. B., & Berkhof, J. (2008). Diagnostic checks for multilevel models. In J. de Leeuw & E. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 141–175). Springer New York. https://doi.org/10.1007/978-0-387-73186-5_3

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (2nd ed.). Sage.

Stegmueller, D. (2013). How many mountries for multilevel modeling? A comparison of frequentist and Bayesian approaches. *American Journal of Political Science*, *57*(3), 748–761. https://doi.org/10.1111/ajps.12001

Sterman, J. D. (2006). Learning from Evidence in a Complex World. *American Journal of Public Health*, *96*(3), 505–514. https://doi.org/10.2105/AJPH.2005.066043

Stockard, J., Gray, J. A., O'Brien, R., & Stone, J. (2009). Race differences in cohort effects on non-marital fertility in the United States. *Social Forces*, *87*(3), 1449–1479. JSTOR.

Suzuki, E. (2012). Time changes, so do people. *Social Science & Medicine*, *75*(3), 452–456. https://doi.org/10.1016/j.socscimed.2012.03.036

Twenge, J. M., Campbell, W. K., & Carter, N. T. (2014). Declines in trust in others and confidence in institutions among American adults and late adolescents, 1972–2012. *Psychological Science*, *25*(10), 1914–1923. https://doi.org/10.1177/0956797614545133

Twenge, J. M., Campbell, W. K., & Sherman, R. A. (2019). Declines in vocabulary among American adults within levels of educational attainment, 1974–2016. *Intelligence*, *76*, 101377. https://doi.org/10.1016/j.intell.2019.101377

Twenge, J. M., Carter, N. T., & Campbell, W. K. (2015). Time Period, Generational, and Age Differences in Tolerance for Controversial Beliefs and Lifestyles in the United States, 1972–2012. *Social Forces*, *94*(1), 379–399. https://doi.org/10.1093/sf/sov050

Twenge, J. M., Sherman, R. A., Exline, J. J., & Grubbs, J. B. (2016). Declines in American adults' religious participation and beliefs, 1972-2014. *SAGE Open*, *6*(1), 2158244016638133. https://doi.org/10.1177/2158244016638133

Twenge, J. M., Sherman, R. A., & Lyubomirsky, S. (2016). More happiness for young people and less for mature adults: Time period differences in subjective well-being in the United States, 1972–2014. *Social Psychological and Personality Science*, *7*(2), 131–141. https://doi.org/10.1177/1948550615602933

Twenge, J. M., Sherman, R. A., & Wells, B. E. (2015). Changes in American Adults' Sexual Behavior and Attitudes, 1972–2012. *Archives of Sexual Behavior*, *44*(8), 2273–2285. https://doi.org/10.1007/s10508-015-0540-2

Twenge, J. M., Sherman, R. A., & Wells, B. E. (2016). Changes in American Adults' Reported Same-Sex Sexual Experiences and Attitudes, 1973–2014. *Archives of Sexual Behavior*, *45*(7), 1713–1730. https://doi.org/10.1007/s10508-016-0769-4

Twenge, J. M., Sherman, R. A., & Wells, B. E. (2017). Sexual inactivity during young adulthood is more common among U.S. Millennials and iGen: Age, period, and cohort effects on having no sexual partners after age 18. *Archives of Sexual Behavior*, *46*(2), 433–440. https://doi.org/10.1007/s10508-016-0798-z

Venzon, D. J., & Moolgavkar, S. H. (1988). A Method for Computing Profile-Likelihood-Based Confidence Intervals. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *37*(1), 87–94. https://doi.org/10.2307/2347496

Wilkes, R. (2011). Re-thinking the decline in trust: A comparison of black and white Americans. *Social Science Research*, *40*(6), 1596–1610. https://doi.org/10.1016/j.ssresearch.2011.06.007

Wilkes, R., & Corrigall-Brown, C. (2011). Explaining time trends in public opinion: Attitudes towards immigration and immigrants. *International Journal of Comparative Sociology*, *52*(1–2), 79–99. https://doi.org/10.1177/0020715210379460

Winship, C., & Harding, D. J. (2008). A mechanism-based approach to the identification of age–period–cohort models. *Sociological Methods & Research*, *36*(3), 362–401. https://doi.org/10.1177/0049124107310635

Woltman, H., Feldstain, A., MacKay, J. C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 52–69. https://doi.org/10.20982/tqmp.08.1.p052

Wu, W., West, S. G., & Taylor, A. B. (2009). Evaluating model fit for growth curve models: Integration of fit indices from SEM and MLM frameworks. *Psychological Methods*, *14*(3), 183–201. https://doi.org/10.1037/a0015858

Yang, Y. (2006). Bayesian inference for hierarchical age-period-cohort models of repeated cross-section survey data. *Sociological Methodology*, *36*(1), 39–74. https://doi.org/10.1111/j.1467-9531.2006.00174.x

Yang, Y. (2008). Social inequalities in happiness in the United States, 1972 to 2004: An age-period-cohort analysis. *American Sociological Review*, *73*(2), 204–226. https://doi.org/10.1177/000312240807300202

Yang, Y. (2011). Chapter 2—Aging, cohorts, and methods. In R. H. Binstock & L. K. George (Eds.), *Handbook of Aging and the Social Sciences (Seventh Edition)* (pp. 17–30). Academic Press. https://doi.org/10.1016/B978-0-12-380880-6.00002-2

Yang, Y., Fu, W. J., & Land, K. C. (2004). A methodological comparison of age-period-cohort models: The intrinsic estimator and conventional generalized linear models. *Sociological Methodology*, *34*(1), 75–110. https://doi.org/10.1111/j.0081-1750.2004.00148.x

Yang, Y., & Land, K. (2013). *Age-period-cohort analysis: New models, methods, and empirical applications*. CLC Press.

Yang, Y., & Land, K. C. (2006). A mixed models approach to the age-period-cohort analysis of repeated cross-section surveys, with an application to data on trends in verbal test scores. *Sociological Methodology*, *36*(1), 75–97. https://doi.org/10.1111/j.1467-9531.2006.00175.x

Yang, Y., & Land, K. C. (2008). Age–period–cohort analysis of repeated cross-section surveys: Fixed or random effects? *Sociological Methods & Research*, *36*(3), 297–326. https://doi.org/10.1177/0049124106292360

Ye, F., & Daniel, L. (2017). The impact of inappropriate modeling of cross-classified data structures on random-slope models. *Journal of Modern Applied Statistical Methods*, *16*(2), 458–484. https://doi.org/10.22237/jmasm/1509495900

Zhang, L. (2017). An age–period–cohort analysis of religious involvement and adult self-rated health: Results from the USA, 1972–2008. *Journal of Religion and Health*, *56*(3), 916–945. https://doi.org/10.1007/s10943-016-0292-x

Zheng, H. (2015). Losing confidence in medicine in an era of medical expansion? *Social Science Research*, *52*, 701–715. https://doi.org/10.1016/j.ssresearch.2014.10.009

Zheng, H., Yang, Y., & Land, K. C. (2011). Variance function regression in hierarchical age-period-cohort models: Applications to the study of self-reported health. *American Sociological Review*, *76*(6), 955–983. https://doi.org/10.1177/0003122411430940

APPENDICES

Appendix A. Compendium of Applied Studies using Hierarchical Age-Period-Cohort Analysis

Author(s)	Topic	Year Range, # of Periods, and Data Source	Cohort Size(s) Used	Justification for Cohort Sizes
An et al. (2021)	Utilization of multiple modes of transportation	2001 – 2017 17 periods National Travel Survey in England	5-year	None
Anderson et al. (2017)	Attitudes toward death penalty	1974 – 2014 27 periods General Social Survey	3-year, 5-year, 7-year	"We also examined models with dummy variables for 3-year and 5-year age groups with each combination of 3-,5-, and 7-year cohorts and 1-2-, and 3-year periods… We conclude from these models that our findings are not sensitive to our choice of coding for age, period, and cohort" (p. 847).
Attell (2020)	Euthanasia and suicide for terminally ill persons	1977 – 2016 23 periods General Social Survey	5-year	None
Bardo (2017)	Happiness	1973 – 1994 19 periods General Social Survey	Not specified	None
Bardo et al. (2017)	Happiness (replication of a different APC study)	1972 – 2014 Number of periods not reported	5-year	"The Baby Boomer cohort is generally defined by demographers to include all persons born between 1946 and

		General Social Survey		1964, but the 5-year birth cohorts that TSL constructed include an additional birth year (i.e, 1945). To make their and out studies comparable, we construct birth cohorts in the same fashion as TSL." (p. 6).
Caren et al. (2011)	Attendance at protests and signing of petitions among Americans	1973 – 2008 8 periods[a] 21 various data sources	4-year	"We group birth cohort into four-year categories. While our groups could have been theoretically driven…no consensus exists on which dates differentiate these groupings…Results from other period and cohort binning strategies provide similar results and are available from the authors" (p, 135).
Carlisle and Clark (2018)	Religion and Environmentalism	1973 – 2014 29 periods General Social Survey	5-year	"the definition of the width of the time intervals is somewhat arbitrary, but the use of 5-year birth cohorts is the norm in age-period-cohort analyses" (p. 236).
Clark (2017)	Gender differences in attitudes toward various social issues	1972 – 2012 Periods: 12, 18, 19, 20, 24, 28[b] General Social Survey	5-year	"the definition of the width of the time intervals is somewhat arbitrary, but the use of five-year birth cohorts is the norm in age-period-cohort analyses" (p. 34).
Clark et al. (2019)	Environmental attitudes	1973 – 2014 28 periods General Social Survey	5-year	"the definition of the width of the time intervals is somewhat arbitrary, but the use of five-year birth cohorts is the norm in age-period-cohort analyses" (p. 22).
Delaruelle et al. (2015)	Self-rated health	2002 – 2012	3-year	None

		8 periods[c] European Social Survey		
Donnelly et al. (2015)	Attitudes towards women's work	1976 – 2013 18 periods General Social Survey	5-year	"We believe that separating the data into 5-year intervals provides the best compromise between specificity and breadth" (p. 45).
Eisenstein et al. (2017)	Political tolerance and religion	1984 – 2014 18 periods General Social Survey	5-year	None
Elias et al. (2015)	Attitudes toward premarital sex	1975 – 2008 20 periods General Social Survey	5-year	None
Fullerton and Dixon (2010b)	Attitudes toward governmental spending on education, health, and social security	1984 – 2008 17 periods General Social Survey	5-year	"Following demographic and other research (e.g., Yang 2008), we code cohorts into dummy variables based on five-year intervals" (pp. 650-651).
Fullerton and Stern (2013)	Gender differences in political participation	1952 – 2004 11 periods American National Election Studies	10-year	None
Gauchat (2012)	Public trust in science	1974 – 2010 26 periods General Social Survey	Not specified	None
Gorodzeisky and	Anti-immigrant attitudes	2002 – 2014 4 periods[d]	3-year, 5-year	"…we used the conventional practice of a

Semyonov (2018)		European Social Survey		5-year interval (Reither et al., 2015)" (p. 35). "In order to conduct a robustness test, we re-estimated the analysis using a 3-year interval for cohorts" (p. 35).
Horowitz (2015)	Civic participation	1973 – 1994 Number of periods not reported Roper Social and Political Trends Data	5-year	None
Johnson (2021)	Psychological distress and mental health treatment	1997 – 2017 21 periods National Health Interview Survey	5-year	None
Johnson and Schwadel (2019)	Attitudes toward environmental spending	1973 – 2014 29 periods General Social Survey	5-year	"Such coding, and consequently the number of level-2 units, comports with extant empirical age-period-cohort research (e.g., Johnson and Schwadel forthcoming; Schwadel and Garneau 2014; Yang 2008)" (p. 922).
Keyes et al. (2011)	Adolescent marijuana use	1976 – 2007 32 periods Monitoring the Future	Not specified	None
Keyes et al. (2012)	Adolescent alcohol use	1976 – 2007 32 periods	5-year	None

		Monitoring the Future		
Keyes et al. (2019)	Depressive symptoms	1991 – 2018 28 periods Monitoring the Future	1-year	None
Koeber and Oberwittler (2019)	Fear of crime among older adults	2002 – 2016 8 periods[e] European Social Survey	3-year	"Cohorts are birth years grouped in three-year intervals starting in 1919 based on group size considerations" (p. 215).
Kowske et al. (2010)	Work attitudes	1985 – 2009 19 periods Kenexa WorkTrends	Unequal cohorts used based on generational theory	"However, the issue then becomes how to group individuals into generations; when should a generation begin and how long should a generation last? Luckily, many generational theorists have hypothesized generations longer than 1 year…" (p. 269).
Masters (2012)	Race differences in mortality risk	1986 – 2006 19 periods National Health Interview Survey	5-year	"The data were collapsed into five-year age by five-year period by five-year cohort cells for two primary reasons: (1) sparse mortality counts in the black mean's and women's individual-level samples preclude stable age-period-cohort modeling, and (2) the aggregated data structure breaks the linear dependency between age, period, and cohort (Glenn 2005)" (p. 782).
Masters et al. (2012)	Educational differences in mortality risk	1986 – 2004 19 periods	5-year	None

		National Health Interview Survey		
Meuleman (2019)	Attitudes toward income redistribution	1978 – 2014 British Social Attitudes Survey (25 periods) General Social Survey (22 periods)	5-year	None
Pampel (2011a)	Gender egalitarianism	1988 – 2002 3 periods International Social Survey Program	1-year with curvilinear cohort term specified	"The use of cohort quadratic terms with five-year age groups and the two or three survey years eliminates the dependency of cohort on age and period" (p. 676).
Pampel (2016)	Social distribution of tolerant sexual attitudes	1973 – 2014 Periods: 17, 25[f] General Social Survey	1-year with curvilinear cohort term specified	None
Pampel and Hunter (2012)	Environmental spending	1973 – 2008 26 periods General Social Survey	1-year with curvilinear cohort term specified	"Based on Bayesian Information Criterion statistics, two quadratic cohort terms perform better than a set of nine dummy variables for 10-year cohorts, and a set of 16 dummy variables for five-year cohorts" (p. 431).
Reither et al. (2009)	Obesity	1976 – 2002 27 periods National Health Interview Survey	5-year	"In this investigation, we group birth cohorts into 5-year intervals, which are conventional in demography (Yang, 2008b)" (p. 1442).

Rosenberger and Steiner (2019)	Attitudes toward unauthorized immigration	1994 – 2010 7 periods American Public Opinion and Foreign Policy Study	5-year	None
Sanderson et al. (2021)	Attitudes toward immigration	1996 – 2018 General Social Survey (9 periods) American National Election Survey (7 periods)	3-year	None
Schwadel (2010a)	Religious nonaffiliation	1973 – 2016 25 periods General Social Survey	5-year	"As Yang et al. (2008), while the choice of intervals is somewhat arbitrary, five-year birth cohorts are the norm in age-period-cohort analyses" (p. 313).
Schwadel (2010b)	Religious service attendance	1972 – 2006 26 periods General Social Survey	10-year	None
Schwadel (2013a)	Attitudes toward prayer and reading the bible in public schools	1974 – 2010 11 periods[g] General Social Survey	5-year	"Following the norm in age-period-cohort analyses, birth cohorts are coded in 5-year intervals (Yang et al. 2008)" (p. 267).
Schwadel (2013b)	Changing strength of religious affiliation	1974 – 2010 Number of periods not reported General Social Survey	5-year	"Birth cohorts are coded into five-year intervals, as is the norm in age-period-cohort analyses (Yang et al. 2008)" (p. 112).

Schwadel and Garneau (2014)	Changing political tolerance	1974 – 2010 22 periods General Social Survey	5-year	None
Schwadel and Ellison (2017)	Support for legalization of marijuana	1973 – 2014 25 periods General Social Survey	5-year	None
Shi et al. (2020)	Perception of crime as country's most important problem	1960 – 2014 55 periods Most Important Problem Dataset	5-year	None
Shu and Meagher (2018)	Gender attitudes	1977 – 2016 20 periods General Social Survey	Roughly 5-year generational cohorts	"We divide the sample into five generations according to prior research on the distinctive life experiences of age groups in the United States… Demographic research commonly creates cohorts by mechanically dividing samples into five-year intervals, which may separate people who share similar life experiences (Yang 2008). Instead, we divide the cohorts around historical events that likely created distinctive experiences during respondents' formative years" (pp. 1249-1250).
Smets and Neundorf (2014)	Voter turnout	1972-2010 28 periods	16 electoral cohorts defined by years in which a	"The grouping of cohorts is also of concern, as it determines the number of observations on the group-level.

		General Social Survey	respondent became eligible to vote	Distinguishing only a few cohorts would not yield enough variation on the cohort-level, increasing the risk of conducting Type-II errors" (p. 44).
Twenge et al. (2014)	Confidence in various social institutions	1972 – 2012 20 periods General Social Survey	Not specified	None
Twenge, Carter, and Campbell (2015)	Attitudes toward controversial beliefs and lifestyles	1972 – 2012 25 periods General Social Survey	5-year	"Data collected over time can be analyzed in many ways, including grouping by 20-year generation blocks, by decades, or individual year. We felt that separating the data into five-year intervals provided the best compromise between specificity and breadth" (p. 385.)
Twenge, Sherman, and Wells (2015)	Sexual behaviors and attitudes toward sexuality	1972 – 2012 Periods: 14, 19 General Social Survey	10-year	"To focus on the general trends, we grouped birth cohorts by decade with the exception of the first cohort (1883-1889) and the last cohort (1990-1994) as they did not make complete decades by themselves" (p. 2277).
Twenge, Sherman, and Lyubomirsky (2016)	Trends in happiness	1972 – 2014 Number of periods not reported[h] Monitoring the Future and General Social Survey	5-year	"We felt that separating the data into 5-year intervals provided the best compromise between specificity and breadth" (p. 3).
Twenge, Sherman,	Same-sex sexual behaviors and attitudes	1973 – 2014 15 periods	10-year	"To focus on the general trends, we grouped birth cohorts by decade with

and Wells (2016)		General Social Survey		the exception of the first cohort (1883-1889) and the last cohort (1990-1996) as they did not make complete decades by themselves" (p. 1717).
Twenge, Sherman, Exline, and Grubbs (2016)	Religious beliefs and participation in religion	1972 – 2014 Periods: 20, 29, 30 General Social Survey	10-year	None
Twenge et al. (2017)	Sexual inactivity	1972 – 2014 15 periods General Social Survey	10-year	"We grouped people by birth decade as a compromise between breadth and depth. Using a larger span (for example, a 20-year generation) risks losing discriminatory power, and a smaller span (such a 5-year groups) risks low sample size" (p.435).
Twenge, Campbell, and Sherman (2019)	Trends in verbal intelligence	1974 – 2016 22 periods General Social Survey	1-year	None
Wilkes (2011)	Race differences in trust	1972 – 2008 23 periods General Social Survey	5-year	None
Wilkes and Corrigall-Brown (2011)	Attitudes towards immigration and immigrants	1987 – 2008 14 periods Environics National Surveys	5-year	None
Yang (2008)	Inequalities in happiness	1972 – 2004 22 periods	5-year	"Specifically, we can use single years of age, time periods corresponding to

		General Social Survey		years in which the surveys are conducted, and birth cohorts defined by five-year intervals, which are conventional in demography" (p. 210).
Zhang (2017)	Religious involvement and self-rated health	1972 – 2008 24 periods General Social Survey	5-year	None
Zheng (2015)	Confidence in institution of medicine	1973 – 2008 27 periods General Social Survey	5-year	None
Zheng et al. (2011)	Self-rated health	1984 – 2007 24 periods National Health Interview Survey	5-year	None

Notes:

a. Caren et al. (2011) pooled items from a variety of datasets and created unequally spaced period groupings. This approach is not conventional. Most researchers use repeated cross-sections from one dataset and use single-year periods.
b. Clark (2017) analyzed seven different outcomes of the GSS data using separate HAPC models. The outcomes were not all asked in the same set of years, therefore some of the models have differing amounts of periods in the analysis.
c. Given the very small number of periods in the data (8), Delaruelle et al. (2015) nest periods into countries, and create an age by period by country clustering and use MCMC estimation procedures to estimate the APC effects. This type of approach is an extension of that proposed by Yang and Land (2006) and is discussed in Bell and Jones (2014a) and Fairbrother (2014).
d. Similar to Delaruelle et al. (2015), Gorodzeisky and Semyonov (2018) nest periods into countries in the cross-classified data structure.
e. Unlike Delaruelle et al. (2015) and Gorodzeisky and Semyonov (2018) do not extend the cross-classification data structure proposed by Yang and Land (2006) and only use eight periods (i.e., they do not nest periods into countries).
f. Pampel (2016) analyzed four different outcomes. Three of the outcomes had twenty-five periods of data available, while one had seventeen.
g. Schwadel (2013a) takes an unusual approach and arbitrarily collapses survey years with adjacent survey years, resulting in a smaller number of periods compared to other studies using General Social Survey data.
h. Twenge, Sherman, and Lyubormirksy (2016) utilize single-year periods in their HAPC analysis, but then examine trends by grouping the period residuals into five-year averages. This is an uncommon approach and it is unclear the exact number of periods utilized in the analysis.

Appendix B. R Program for Real Data Study and Monte Carlo Simulation

```
#load necessary packages#
library(foreign)
library(dplyr)
library(ggplot2)
library(lme4)
library(lmerTest)
library(parallel)
library(tidyr)
library(openxlsx)
library(cowplot)

################################################################################
################################################################################

#step 1 - real data study#

#set working directory#
setwd("~/Dropbox/Dissertation/data/RDS Data_bothdivorce")

#function to read data dictionary#
read.dct <- function(dct, labels.included = "yes") {
  temp <- readLines(dct)
  temp <- temp[grepl("_column", temp)]
  switch(labels.included,
         yes = {
           pattern <- "_column\\(([0-9]+)\\)\\s+([a-z0-9]+)\\s+(.*)\\s+%([0-9]+)[a-
z]\\s+(.*)"
           classes <- c("numeric", "character", "character", "numeric", "charac-
ter")
           N <- 5
           NAMES <- c("StartPos", "Str", "ColName", "ColWidth", "ColLabel")
         },
         no = {
           pattern <- "_column\\(([0-9]+)\\)\\s+([a-z0-9]+)\\s+(.*)\\s+%([0-9]+).*"
           classes <- c("numeric", "character", "character", "numeric")
           N <- 4
           NAMES <- c("StartPos", "Str", "ColName", "ColWidth")
         })
  temp_metadata <- setNames(lapply(1:N, function(x) {
    out <- gsub(pattern, paste("\\", x, sep = ""), temp)
    out <- gsub("^\\s+|\\s+$", "", out)
    out <- gsub('\"', "", out, fixed = TRUE)
    class(out) <- classes[x] ; out }), NAMES)
  temp_metadata[["ColName"]] <- make.names(gsub("\\s", "",
                                                temp_metadata[["ColName"]]))
  temp_metadata
}

#function to read data#
read.dat <- function(dat, metadata_var, labels.included = "yes") {
  read.fwf(dat, widths = metadata_var[["ColWidth"]],
           col.names = metadata_var[["ColName"]])
}
```

```
#call data and apply formats#
GSS_metadata <- read.dct("GSS.dct")
GSS_ascii <- read.dat("GSS.dat", GSS_metadata)
attr(GSS_ascii, "col.label") <- GSS_metadata[["ColLabel"]]
gss <- GSS_ascii

#make all variable names lowercase#
colnames(gss) <- tolower(colnames(gss))

#limit dataset to years in which all abortion questions were asked#
gss_sub <- gss %>%
  filter(year %in% c(seq(1972,1978), 1980, 1982, 1983, 1984, 1985, 1987,
                     1988, 1989, 1990, 1991, 1993, seq(1994,2016,2)))

#remove observations whose birth cohorts were missing#
gss_sub <- gss_sub %>% filter(cohort != 9999)

#remove individuals who were DK or no answer or not applicable on divlaw#
gss_sub <- gss_sub %>% filter(divlaw %in% c(1,2,3))

#remove individuals who were DK or no answer or not applicable on divorce#
gss_sub <- gss_sub %>% filter(divorce %in% c(1,2))


#ensure there are no non-discrete missing data on sex variable#
length(subset(is.na(gss_sub$sex), TRUE)) == nrow(gss_sub)

#limit dataset to only valid abortion responses#
gss_sub <- gss_sub %>%
  filter(abdefect %in% c(1,2) &
           abhlth %in% c(1,2) &
           abnomore %in% c(1,2) &
           abpoor %in% c(1,2) &
           abrape %in% c(1,2) &
           absingle %in% c(1,2))

#recode each item#
gss_sub$abdefect <- ifelse(gss_sub$abdefect == 2, 0, gss_sub$abdefect)
gss_sub$abhlth <- ifelse(gss_sub$abhlth == 2, 0, gss_sub$abhlth)
gss_sub$abnomore <- ifelse(gss_sub$abnomore == 2, 0, gss_sub$abnomore)
gss_sub$abpoor <- ifelse(gss_sub$abpoor == 2, 0, gss_sub$abpoor)
gss_sub$abrape <- ifelse(gss_sub$abrape == 2, 0, gss_sub$abrape)
gss_sub$absingle <- ifelse(gss_sub$absingle == 2, 0, gss_sub$absingle)

#create abortion index#
gss_sub$abortion <- with(gss_sub, abdefect + abhlth + abnomore + abpoor +
                           abrape + absingle)

#scale the index#
gss_sub$abscaled <- gss_sub$abortion*100

#create necessary model variables#
#indicator for female respondent, 1 = female, 0 = male#
gss_sub <- gss_sub %>%
  mutate(age = year - cohort, #age in years at time of interview#
         age2 = age*age, #age squared term
         female = ifelse(sex == 2, 1, 0),
```

```
          ndivlaw = ifelse(divlaw == 1, 1, 0),
          ndivorce = ifelse(divorce == 1, 1, 0))

#examine cohorts and years that exist in the data#
cohorts <- gss_sub %>% group_by(cohort) %>%
  summarise(freq = n())

years <- gss_sub %>% group_by(year) %>%
  summarise(freq = n())

#calculate 5-year birth cohorts#
gss_sub$cohort5 <- findInterval(gss_sub$cohort, vec = seq(min(gss_sub$cohort),
                                                          max(gss_sub$cohort),
                                                          by = 5))

#create standardized age terms#
gss_sub$agez <- scale(gss_sub$age, center = TRUE, scale = TRUE)
gss_sub$agez2 <- gss_sub$agez^2

#examine age distribution#
ggplot(data = gss_sub, aes(x = age)) + theme_bw() +
  geom_histogram(binwidth = 0.5) + labs(x = "Age", y = "Frequency")

gss_age <- gss_sub %>% group_by(age) %>% summarise(freq = n()) %>%
  mutate(age_prop = freq / nrow(gss_sub))

#examine sex distribution#
gss_sub %>% group_by(female) %>% summarise(prop = n()/nrow(gss_sub))

#create cohort level divorce variable#
divorce_cohort5 <- gss_sub %>%
  group_by(cohort5) %>%
  summarise(mean_divorce = mean(ndivorce)*100)

gss_sub <- gss_sub %>%
  left_join(divorce_cohort5, by = "cohort5")

#histogram at cohort level#
hist(divorce_cohort5$mean_divorce)

#histogram at person level#
hist(gss_sub$ndivorce)
gss_sub %>% group_by(ndivorce) %>% summarise(prop = n()/nrow(gss_sub))

#center the variable#
gss_sub$divorce_cent <- gss_sub$mean_divorce -
  mean(divorce_cohort5$mean_divorce)

#create divlaw variable at period level#
ndivlaw_year <- gss_sub %>%
  group_by(year) %>%
  summarise(ndivlaw_year = mean(ndivlaw)*100)

gss_sub <- gss_sub %>%
  left_join(ndivlaw_year, by = "year")
```

```
#histogram at period level#
hist(ndivlaw_year$ndivlaw_year)

#histogram at person level#
hist(gss_sub$ndivlaw)

#center the variable#
gss_sub$ndivlaw_cent <- gss_sub$ndivlaw_year - mean(ndivlaw_year$ndivlaw_year)


#table 1 - descriptive statistics#
rbind(gss_sub %>%
        summarize(mean = mean(abscaled), sd = sd(abscaled), min = min(abscaled),
                  max = max(abscaled)) %>%
        mutate(par = "Abortion"),
      gss_sub %>%
        summarize(mean = mean(age), sd = sd(age), min = min(age),
                  max = max(age)) %>%
        mutate(par = "Age"),
      gss_sub %>%
        summarize(mean = mean(female), sd = sd(female), min = min(female),
                  max = max(female)) %>%
        mutate(par = "Female"),
      ndivlaw_year %>%
        summarize(mean = mean(ndivlaw_year), sd = sd(ndivlaw_year),
                  min = min(ndivlaw_year), max = max(ndivlaw_year)) %>%
        mutate(par = "Divorce Law"),
      divorce_cohort5 %>%
        summarize(mean = mean(mean_divorce), sd = sd(mean_divorce),
                  min = min(mean_divorce), max = max(mean_divorce)) %>%
        mutate(par = "Divorce"))

#baseline and conditional models w/ outcome scaled and predictors centered#
mod_5bsc <- lmer(abscaled ~ agez + agez2 + (1|cohort5) + (1|year),
                 data = gss_sub, REML = TRUE)

summary(mod_5bsc)

  #variance components#
  VarCorr(mod_5bsc)

mod_5csc <- lmer(abscaled ~ agez + agez2 + female + divorce_cent +
                   ndivlaw_cent + (1|cohort5) + (1|year),
                 data = gss_sub, REML = TRUE)

summary(mod_5csc)

  #variance components#
  VarCorr(mod_5csc)

################################################################################
################################################################################
#step 2 - monte carlo simulation#

#set replications for each study condition#
reps <- 1000
```

```
#generate study conditions#
years <- c("10", "20", "30")
csm_dgm <- c("3-Year", "5-Year", "10-Year")
per_var <- c("low", "high")
coh_var <- c("low", "high")

conditions <- expand.grid(years, csm_dgm, per_var, coh_var) %>%
  select(years = Var1, csm_dgm = Var2, per_var = Var3, coh_var = Var4) %>%
  #set proportion of variability for periods and cohorts#
  mutate(per_varn = ifelse(per_var == "low", 0.02, 0.10),
         coh_varn = ifelse(coh_var == "low", 0.02, 0.10),
         #set total variability for baseline and conditional models#
         total_b = 40745.70,
         total_c = 40474.60)

#use variance components to reallocate variability across conditions#
conditions <- conditions %>%
  mutate(u0j_b = coh_varn * total_b,
         v0k_b = per_varn * total_b,
         eijk_b = total_b - u0j_b - v0k_b,
         u0j_c = coh_varn * total_c,
         v0k_c = per_varn * total_c,
         eijk_c = total_c - u0j_c - v0k_c,
         condition = seq(1, 36, 1)) %>% #set condition number#
  select(condition, years, csm_dgm, per_var, coh_var, u0j_b, v0k_b, eijk_b,
         u0j_c, v0k_c, eijk_c)

#list of conditions for parallel processing#
nrep <- rep((1:reps), each = nrow(conditions))
con_combined<- cbind(nrep, do.call(rbind, replicate(n = reps, conditions,
                                                    simplify = FALSE)))
con_combined <- con_combined[order(con_combined$condition),]
con_combined$seed <- seq(1,nrow(con_combined),1)
simparms <- split(con_combined, 1:nrow(con_combined))
remove(reps, nrep, con_combined, years, coh_var, csm_dgm, per_var)

##############################################################################
runCondition <- function(myParams){
  library(dplyr)
  library(stringr)
  library(feather)
  library(purrr)
  library(betafunctions)

  #set seed - use padding for purposes of exporting datasets below#
  n_seed <- str_pad(as.character(myParams$seed), 5, pad = "0")
  set.seed(n_seed)

  #set number of periods#
  n_years <- myParams$years

  #set cohort selection mechanism to use in data generating model#
  csm_dgm <- myParams$csm_dgm

  #set number of observations and survey years#
  if(n_years == "10"){
```

```
  n_obs <- 15000
  years <- c(seq(1972,1978,1),1980,1982,1983)
}

if(n_years == "20"){
  n_obs <- 30000
  years <- c(seq(1972,1978,1),1980,seq(1982,1985,1),seq(1987,1991,1),
             1993,1994,1996)
}

if(n_years == "30"){
  n_obs <- 45000
  years <- c(seq(1972,1978,1),1980,seq(1982,1985,1),seq(1987,1991,1),1993,
             seq(1994,2016,2))
}

#set age distribution#
age_dist <- data.frame(age = seq(18,89,1),
                       freq = c(11,24,55,96,115,159,177,215,240,258,298,231,
                                296,274,316,280,294,266,274,269,282,277,250,
                                246,266,239,225,219,219,209,229,265,221,241,
                                225,246,201,207,228,218,243,201,222,223,210,
                                214,178,194,164,188,168,177,158,155,158,136,
                                133,110,141,121,94,82,81,83,65,49,43,48,37,
                                45,23,62)) %>%
  mutate(age_prop = freq / sum(freq))

#generate age variable#
age <- sample(age_dist$age, size = n_obs, replace = TRUE,
              prob = age_dist$age_prop)

#standardize and square the term#
agez <- (age - mean(age)) / sd(age)
agez2 <- agez*agez

remove(age_dist)

#generate periods and cohorts#
year <- sample(years, size = n_obs, replace = TRUE)
cohort <- year - age

#generate cohort groupings for generating and estimating models#
cgroup3 <- findInterval(cohort, vec = seq(min(cohort), max(cohort), by = 3))
cgroup5 <- findInterval(cohort, vec = seq(min(cohort), max(cohort), by = 5))
cgroup10 <- findInterval(cohort, vec = seq(min(cohort), max(cohort), by = 10))

#set cohort groupings to use in data generating models#
if(csm_dgm == "3-Year"){
  cgroup_dgm <- cgroup3
}

if(csm_dgm == "5-Year"){
  cgroup_dgm <- cgroup5
}

if(csm_dgm == "10-Year"){
  cgroup_dgm <- cgroup10
```

```
}
#generate variability in cohorts#
c_varb <- data.frame(cgroup_dgm = unique(cgroup_dgm),
                     c_varb = rnorm(n = length(unique(cgroup_dgm)),
                                    mean = 0, sd = sqrt(myParams$u0j_b)))

c_varc <- data.frame(cgroup_dgm = unique(cgroup_dgm),
                     c_varc = rnorm(n = length(unique(cgroup_dgm)),
                                    mean = 0, sd = sqrt(myParams$u0j_c)))

#generate variability in periods#
y_varb <- data.frame(year = unique(year),
                     y_varb = rnorm(n = length(unique(year)),
                                    mean = 0, sd = sqrt(myParams$v0k_b)))

y_varc <- data.frame(year = unique(year),
                     y_varc = rnorm(n = length(unique(year)),
                                    mean = 0, sd = sqrt(myParams$v0k_c)))

#generate level-one variability#
ind_varb <- rnorm(n = n_obs, mean = 0, sd = sqrt(myParams$eijk_b))

ind_varc <- rnorm(n = n_obs, mean = 0, sd = sqrt(myParams$eijk_c))

#generate indicator for female#
female <- rbernoulli(n = n_obs, p = 0.559)

#create cohort-level ever divorced for data estimating models#
l1div <- rbernoulli(n = n_obs, p = 0.199)

div_c3 <- data.frame(cgroup3, l1div) %>%
  group_by(cgroup3) %>%
  summarise(div3 = mean(l1div)*100)

div_c3$div3_cent <- div_c3$div3 - mean(div_c3$div3)

div_c5 <- data.frame(cgroup5, l1div) %>%
  group_by(cgroup5) %>%
  summarise(div5 = mean(l1div)*100)

div_c5$div5_cent <- div_c5$div5 - mean(div_c5$div5)

div_c10 <- data.frame(cgroup10, l1div) %>%
  group_by(cgroup10) %>%
  summarise(div10 = mean(l1div)*100)

div_c10$div10_cent <- div_c10$div10 - mean(div_c10$div10)

#set cohort-level ever divorce for use in data generating models#
#note that this generates with the grand mean centered divorce variable#
if(csm_dgm == "3-Year"){
  div_coh <- div_c3 %>%
    select(cgroup_dgm = cgroup3, divorce = div3_cent)
}

if(csm_dgm == "5-Year"){
  div_coh <- div_c5 %>%
```

```
      select(cgroup_dgm = cgroup5, divorce = div5_cent)
  }

  if(csm_dgm == "10-Year"){
    div_coh <- div_c10 %>%
      select(cgroup_dgm = cgroup10, divorce = div10_cent)
  }


  #create grand mean centered period level divorce law attitude#
  divlaw_per <- data.frame(year = years,
                           divlaw_dist = rBeta.4P(n = length(years),
                                                  l = 18, u = 30,
                                                  alpha = 3, beta = 8))

  divlaw_per$divlaw <- with(divlaw_per, divlaw_dist - mean(divlaw_dist))

  #combine generated data into dataframe#
  d <- data.frame(cbind(year, age, agez, agez2, cohort, cgroup3, cgroup5,
                        cgroup10, cgroup_dgm, ind_varb, ind_varc, female)) %>%
    #link cohort variability#
    left_join(c_varb, by = "cgroup_dgm") %>%
    left_join(c_varc, by = "cgroup_dgm") %>%
    #link period variability#
    left_join(y_varb, by = "year") %>%
    left_join(y_varc, by = "year") %>%
    #link level-two predictors#
    left_join(div_coh, by = "cgroup_dgm") %>%
    left_join(divlaw_per, by = "year") %>%
    #link different cohort versions of divorce for estimating models#
    left_join(div_c3, by = "cgroup3") %>%
    left_join(div_c5, by = "cgroup5") %>%
    left_join(div_c10, by = "cgroup10")

  remove(year, age, agez, agez2, cohort, cgroup3, cgroup5, cgroup10, ind_varb,
         ind_varc, c_varb, y_varb, c_varc, y_varc, female, div_coh, divlaw_per,
         div_c3, div_c5, div_c10)

  #generate outcome#
  d$abortb <- 378.373 - 4.793*d$agez - 7.754*d$agez2 + d$c_varb + d$y_varb +
    d$ind_varb

  d$abortc <- 373.057 - 5.9251*d$agez - 5.737*d$agez2 - 11.0714*d$female +
    3.6598*d$divorce + 3.3346*d$divlaw + d$c_varc + d$y_varc + d$ind_varc

  d <- d %>%
    select(year, age, agez, agez2, cohort, cgroup3, cgroup5, cgroup10, abortb,
           abortc, cgroup_dgm, female, divlaw, div3_cent, div5_cent, div10_cent)

  #save dataset#
  setwd("~/Desktop/dgm_dat2")
  write_feather(d, path = paste("dat_", n_seed,
                                ".feather", sep = ""))
  remove(d)

}
```

```
################################################################################
#run data generating models#
start_time <- Sys.time()
cl <- makeCluster(mc <- getOption("cl.cores", as.numeric(detectCores())))
parLapply(cl, simparms, runCondition)
stopCluster(cl)
end_time <- Sys.time()
time_taken <- end_time - start_time
time_taken

################################################################################

#function for estimating models#
est_mods <- function(filename){
  library(lme4)
  library(feather)
  library(broom.mixed)
  library(stringr)
  library(dplyr)

  #call data#
  dat <- read_feather(filename)

  #estimate baseline models#
  mod3b <- lmer(abortb ~ agez + agez2 + (1|cgroup3) + (1|year),
                data = dat, REML = TRUE)

  mod5b <- lmer(abortb ~ agez + agez2 + (1|cgroup5) + (1|year),
                data = dat, REML = TRUE)

  mod10b <- lmer(abortb ~ agez + agez2 + (1|cgroup10) + (1|year),
                 data = dat, REML = TRUE)

  #function for corrected AIC#
  aicc <- function(model, mod_name, n){
    aicc <- (((-2*logLik(model)) + (2*6*n)) / (n - 6 - 1))[1]
    res <- data.frame(fit = aicc, model = mod_name, index = "AICC")
    return(res)
  }

  aicc_fit <- rbind(aicc(model = mod3b, mod_name = "cohort3", n = nrow(dat)),
                    aicc(model = mod5b, mod_name = "cohort5", n = nrow(dat)),
                    aicc(model = mod10b, mod_name = "cohort10", n = nrow(dat)))

  #function for corrected BIC#
  bicc <- function(model, mod_name, cgroup){
    bicc <- ((-2*logLik(model)) +
               (log((length(unique(dat$year)))*(length(unique(cgroup))))*6))[1]
    res <- data.frame(fit = bicc, model = mod_name, index = "BICC")
  }

  bicc_fit <- rbind(bicc(model = mod3b, mod_name = "cohort3",
                         cgroup = dat$cgroup3),
                    bicc(model = mod5b, mod_name = "cohort5",
                         cgroup = dat$cgroup5),
                    bicc(model = mod10b, mod_name = "cohort10",
                         cgroup = dat$cgroup10))
```

```
  #save model fit indices for baseline models#
  mfit <- rbind((AIC(mod3b, mod5b, mod10b) %>%
                   mutate(model = c("cohort3","cohort5","cohort10"),
                          index = "AIC") %>%
                   select(fit = AIC, model, index)),
                (BIC(mod3b, mod5b, mod10b) %>%
                   mutate(model = c("cohort3","cohort5","cohort10"),
                          index = "BIC") %>%
                   select(fit = BIC, model, index)),
                aicc_fit,
                bicc_fit)

  #function to flag models that failed to converge#
  convergence <- function(mod){
    con <- ifelse("TRUE" %in%
                    str_detect(mod@optinfo$conv$lme4$messages,
                               "failed to converge"), 1, 0)
    return(con)
  }

  #convergence status for baseline models#
  convergeb <- data.frame(rbind(convergence(mod = mod3b),
                                convergence(mod = mod5b),
                                convergence(mod = mod10b)))
  convergeb$em <- c("cgroup3", "cgroup5", "cgroup10")
  colnames(convergeb) <- c("failed_to_converge", "em")

  #estimate conditional models#
  mod3c <- lmer(abortc ~ agez + agez2 + female + div3_cent + divlaw +
                  (1|cgroup3) + (1|year), data = dat, REML = TRUE)

  mod5c <- lmer(abortc ~ agez + agez2 + female + div5_cent + divlaw +
                  (1|cgroup5) + (1|year), data = dat, REML = TRUE)

  mod10c <- lmer(abortc ~ agez + agez2 + female + div10_cent + divlaw +
                   (1|cgroup10) + (1|year), data = dat, REML = TRUE)

  #convergence status for conditional models#
  convergec <- data.frame(rbind(convergence(mod = mod3c),
                                convergence(mod = mod5c),
                                convergence(mod = mod10c)))
  convergec$em <- c("cgroup3", "cgroup5", "cgroup10")
  colnames(convergec) <- c("failed_to_converge", "em")

  #save results#
  res <- list(fix_eff3 = tidy(mod3c, conf.int = TRUE),
              fix_eff5 = tidy(mod5c, conf.int = TRUE),
              fix_eff10 = tidy(mod10c, conf.int = TRUE),
              model_fit = mfit,
              converge_statb = convergeb,
              converge_statc = convergec)

  return(res)
  remove(res, mod3b, mod5b, mod10b, mod3c, mod5c, mod10c, convergeb, convergec,
         aicc_fit,bicc_fit)
}
```

```r
#call generated datasets and separate into conditions#
filenames <- list.files(path = "~/Desktop/dgm_dat2", full.names = TRUE)
con1 <- filenames[1:1000]
con2 <- filenames[1001:2000]
con3 <- filenames[2001:3000]
con4 <- filenames[3001:4000]
con5 <- filenames[4001:5000]
con6 <- filenames[5001:6000]
con7 <- filenames[6001:7000]
con8 <- filenames[7001:8000]
con9 <- filenames[8001:9000]
con10 <- filenames[9001:10000]
con11 <- filenames[10001:11000]
con12 <- filenames[11001:12000]
con13 <- filenames[12001:13000]
con14 <- filenames[13001:14000]
con15 <- filenames[14001:15000]
con16 <- filenames[15001:16000]
con17 <- filenames[16001:17000]
con18 <- filenames[17001:18000]
con19 <- filenames[18001:19000]
con20 <- filenames[19001:20000]
con21 <- filenames[20001:21000]
con22 <- filenames[21001:22000]
con23 <- filenames[22001:23000]
con24 <- filenames[23001:24000]
con25 <- filenames[24001:25000]
con26 <- filenames[25001:26000]
con27 <- filenames[26001:27000]
con28 <- filenames[27001:28000]
con29 <- filenames[28001:29000]
con30 <- filenames[29001:30000]
con31 <- filenames[30001:31000]
con32 <- filenames[31001:32000]
con33 <- filenames[32001:33000]
con34 <- filenames[33001:34000]
con35 <- filenames[34001:35000]
con36 <- filenames[35001:36000]

#function to run estimating models#
run_emods <- function(condition){
  start_time <- Sys.time()
  cl <- makeCluster(mc <- getOption("cl.cores", as.numeric(detectCores())))
  simulationResults <- parLapply(cl, condition, est_mods)
  stopCluster(cl)
  end_time <- Sys.time()
  time_taken <- end_time - start_time
  print(time_taken)

  return(simulationResults)
}

#run estimating models by condition#
res_con1 <- run_emods(condition = con1)
res_con2 <- run_emods(condition = con2)
res_con3 <- run_emods(condition = con3)
```

```
res_con4 <- run_emods(condition = con4)
res_con5 <- run_emods(condition = con5)
res_con6 <- run_emods(condition = con6)
res_con7 <- run_emods(condition = con7)
res_con8 <- run_emods(condition = con8)
res_con9 <- run_emods(condition = con9)
res_con10 <- run_emods(condition = con10)
res_con11 <- run_emods(condition = con11)
res_con12 <- run_emods(condition = con12)
res_con13 <- run_emods(condition = con13)
res_con14 <- run_emods(condition = con14)
res_con15 <- run_emods(condition = con15)
res_con16 <- run_emods(condition = con16)
res_con17 <- run_emods(condition = con17)
res_con18 <- run_emods(condition = con18)
res_con19 <- run_emods(condition = con19)
res_con20 <- run_emods(condition = con20)
res_con21 <- run_emods(condition = con21)
res_con22 <- run_emods(condition = con22)
res_con23 <- run_emods(condition = con23)
res_con24 <- run_emods(condition = con24)
res_con25 <- run_emods(condition = con25)
res_con26 <- run_emods(condition = con26)
res_con27 <- run_emods(condition = con27)
res_con28 <- run_emods(condition = con28)
res_con29 <- run_emods(condition = con29)
res_con30 <- run_emods(condition = con30)
res_con31 <- run_emods(condition = con31)
res_con32 <- run_emods(condition = con32)
res_con33 <- run_emods(condition = con33)
res_con34 <- run_emods(condition = con34)
res_con35 <- run_emods(condition = con35)
res_con36 <- run_emods(condition = con36)

simulationResults <- c(res_con1, res_con2, res_con3, res_con4, res_con5,
                       res_con6, res_con7, res_con8, res_con9, res_con10,
                       res_con11, res_con12, res_con13, res_con14, res_con15,
                       res_con16, res_con17, res_con18, res_con19, res_con20,
                       res_con21, res_con22, res_con23, res_con24, res_con25,
                       res_con26, res_con27, res_con28, res_con29, res_con30,
                       res_con31, res_con32, res_con33, res_con34, res_con35,
                       res_con36)

rm(list = ls(pattern = "^res_con"))
rm(list = ls(pattern = "^con"))


################################################################################

#outcome 1 - model convergence#
converge <- do.call(rbind,
                    do.call(rbind,
                            lapply(simulationResults, '[', "converge_statb")))

converge$condition <- rep(paste("Condition", seq(1,36)), each = 3000)
converge$model <- rep(seq(1,1000), each = 3, times = 36)
```

```
converge_by_condition <- converge %>%
  group_by(condition, em) %>%
  summarise(failed = sum(failed_to_converge),
            failed_percent = (sum(failed_to_converge)/1000 * 100))

converge_wide <- converge_by_condition %>%
  select(-failed) %>%
  spread(em, failed_percent) %>%
  mutate(con = as.numeric(substr(condition, 11,12))) %>%
  arrange(con) %>%
  select(-con)

conditions$condition <- paste("Condition", conditions$condition)

converge_wide <- converge_wide %>%
  left_join(conditions %>% select(condition, years, csm_dgm, per_var, coh_var),
            by = "condition")


converge_wide <- converge_wide %>%
  select(condition, years, csm_dgm, per_var, coh_var, cgroup3, cgroup5,
         cgroup10)

################################################################################

#outcome 2 - relative absolute bias#

#get model convergence for conditional models#
converge_c <- do.call(rbind,
                      do.call(rbind,
                              lapply(simulationResults, '[', "converge_statc")))

converge_c$condition <- rep(paste("Condition", seq(1,36)), each = 3000)
converge_c$model <- rep(seq(1,1000), each = 3)

#extract fixed effects and variance components#
fix_rpb <- function(result){
  res <- do.call(rbind,
                 do.call(rbind,
                         lapply(simulationResults, `[`, result)))

  res$condition <- rep(paste("Condition", seq(1,36)), each = 9000)

  res$result <- ifelse(res$effect == "ran_pars",
                       res$estimate^2, res$estimate)

  res$par <- ifelse(res$effect == "ran_pars", res$group,
                    res$term)

  res$model <- rep(seq(1,1000), each = 9, times = 36)

  return(res)
}

fix_eff3 <- fix_rpb(result = "fix_eff3")
fix_eff5 <- fix_rpb(result = "fix_eff5")
fix_eff10 <- fix_rpb(result = "fix_eff10")
```

```
#add indicator for estimating model#
fix_eff3$em <- "cgroup3"
fix_eff5$em <- "cgroup5"
fix_eff10$em <- "cgroup10"

#combine fixed effects results#
fix_eff <- rbind(fix_eff3, fix_eff5, fix_eff10)

#remove models that did not converge#
fix_eff <- fix_eff %>%
  left_join(converge_c, by = c("condition", "model", "em")) %>%
  filter(failed_to_converge == 0)

#set divorce variable name to be consistent across estimating models#
fix_eff$par <- ifelse(fix_eff$par %in% c("div3_cent", "div5_cent",
                                         "div10_cent"), "divorce", fix_eff$par)

#set true parameter values#
truth <- conditions %>%
  select(condition, csm_dgm, Residual = eijk_c, year = v0k_c,
         cgroup = u0j_c) %>%
  mutate('(Intercept)' = 373.057,
         agez = -5.9251,
         agez2 = -5.737,
         female = -11.0714,
         divorce = 3.6598,
         divlaw = 3.3346) %>%
  gather(par, truth, Residual:divlaw)

#update parameter names of cohorts to match returned objects from EMs#
truth$par <- ifelse(truth$par == "cgroup",
                    paste("cgroup", substr(truth$csm_dgm,1,1), sep = ""),
                    truth$par)

truth$par <- ifelse(truth$par == "cgroup1", "cgroup10", truth$par)


#calculate relative absolute bias#
rpb <- fix_eff %>%
  group_by(condition, em, par) %>%
  summarise(avg = mean(result)) %>%
  left_join(truth, by = c("condition", "par"))

rpb <- rpb %>%
  left_join(conditions %>% select(condition, u0j_c), by = "condition")

rpb$truth <- ifelse(is.na(rpb$truth), rpb$u0j_c, rpb$truth)

rpb <- rpb %>%
  mutate(rpb = ((avg - truth)/abs(truth))*100) %>%
  select(condition, em, par, avg, truth, rpb)

#round off results for table#
rpb$rpb_round <- round(rpb$rpb, 3)

#rename cohort variance components to all be the same for function below#
```

```
rpb$par <- ifelse(substr(rpb$par,1,6) == "cgroup", "cgroup", rpb$par)

#for each parameter, gather results in wide format by condition#
rpb_gather <- function(parameter){
  res <- rpb %>%
    filter(par == parameter) %>%
    select(condition, em, rpb_round) %>%
    spread(em, rpb_round) %>%
    mutate(connum = as.numeric(substr(condition, 11,13))) %>%
    left_join(conditions %>% select(condition, years, csm_dgm, per_var,
                                    coh_var), by = "condition") %>%
    arrange(connum) %>%
    select(condition, years, csm_dgm, per_var, coh_var, cgroup3, cgroup5,
           cgroup10)

  return(res)
}

rpb_int <- rpb_gather(parameter = "(Intercept)")
rpb_age <- rpb_gather(parameter = "agez")
rpb_age2 <- rpb_gather(parameter = "agez2")
rpb_female <- rpb_gather(parameter = "female")
rpb_divorce <- rpb_gather(parameter = "divorce")
rpb_divlaw <- rpb_gather(parameter = "divlaw")
rpb_l1res <- rpb_gather(parameter = "Residual")
rpb_perres <- rpb_gather(parameter = "year")
rpb_cohres <- rpb_gather(parameter = "cgroup")

################################################################################

#outcome 3 - coverage rates#

#extract confidence intervals for fixed effects#
#this draws on the fixed effects estimates already obtained for outcome 2#
#and already has models that failed to converge removed#

#start by removing variance components from results#
conf_fix <- fix_eff %>%
  filter(is.na(std.error) == FALSE)

#link true parameter values#
conf_fix <- conf_fix %>%
  left_join(truth, by = c("condition", "par"))

#determine if confidence interval contains truth#
conf_fix$int_con <- ifelse(conf_fix$conf.low <= conf_fix$truth &
                             conf_fix$conf.high >= conf_fix$truth, 1, 0)

#sum up by condition and estimating model#
cov_rate <- conf_fix %>%
  group_by(condition, par, em) %>%
  summarise(tot_cov = sum(int_con))

#get denominator for number of models converged by condition and est. model#
cov_denom <- converge_c %>%
  group_by(condition, em) %>%
  summarise(tot_failed = sum(failed_to_converge),
```

```
            tot_converge = 1000 - tot_failed)

#link denominator to numerators for coverage rate#
cov_rate <- cov_rate %>%
  left_join(cov_denom %>% select(-tot_failed), by = c("condition", "em"))

#calculate coverage rate#
cov_rate$cov_rate <- (cov_rate$tot_cov / cov_rate$tot_converge)*100

#for each parameter, gather results in wide format by condition#
cov_gather <- function(parameter){
  res <- cov_rate %>%
    ungroup() %>%
    filter(par == parameter) %>%
    select(condition, em, cov_rate) %>%
    spread(em, cov_rate) %>%
    mutate(connum = as.numeric(substr(condition, 11,13))) %>%
    left_join(conditions %>% select(condition, years, csm_dgm, per_var,
                                    coh_var), by = "condition") %>%
    arrange(connum) %>%
    select(condition, years, csm_dgm, per_var, coh_var, cgroup3, cgroup5,
           cgroup10)

  return(res)
}

cov_int <- cov_gather(parameter = "(Intercept)")
cov_age <- cov_gather(parameter = "agez")
cov_age2 <- cov_gather(parameter = "agez2")
cov_female <- cov_gather(parameter = "female")
cov_divorce <- cov_gather(parameter = "divorce")
cov_divlaw <- cov_gather(parameter = "divlaw")


################################################################################

#outcome 4 - model fit indices#

#for each set of estimating models find those where all 3 converged#
conv_agg <- converge %>%
  group_by(condition, model) %>%
  summarise(failed_converge = sum(failed_to_converge)) %>%
  filter(failed_converge == 0) %>%
  mutate(keep = 1)

#get denominator to use in proportion of total executed models by condition#
conv_denom <- conv_agg %>%
  group_by(condition) %>%
  summarise(denominator = sum(keep))

#extract model fit results from estimating models#
mfit <- do.call(rbind,
                do.call(rbind,
                        lapply(simulationResults, `[`, "model_fit"))) %>%
  mutate(model = rep(seq(1:1000), each = 12, times = 36),
         em = rep(c("3-Year", "5-Year", "10-Year"), times = 144000))
```

```
mfit$condition <- rep(paste("Condition", seq(1,36)), each = 12000)
#remove replications where all 3 models did not converge#
mfit <- mfit %>%
  left_join(conv_agg %>% select(condition, model, keep),
            by = c("condition", "model"))

mfit <- subset(mfit, keep == 1)

#link which cohort selection mechanism should have been identified#
mfit <- mfit %>%
  left_join(conditions %>% select(condition, csm_dgm),
            by = c("condition"))

mfit$csm_dgm <- as.character(mfit$csm_dgm)

#function to calculate proportion of times the fit index identified the#
#correct model#
mfit_correct <- function(fit_index){
  result <- mfit %>%
    filter(index == fit_index) %>%
    arrange(condition, model, fit) %>%
    group_by(condition, model) %>%
    filter(row_number()==1) %>%
    mutate(pass = ifelse(em == csm_dgm, 1, 0))

  result_con <- result %>%
    group_by(condition) %>%
    summarise(numerator = sum(pass)) %>%
    left_join(conv_denom, by = "condition") %>%
    mutate(prop = (numerator/denominator)*100) %>%
    select(condition, prop)

  result_con$condition <- as.numeric(substr(result_con$condition, 11,12))
  result_con <- result_con %>% arrange(condition)
}

aic <- mfit_correct(fit_index = "AIC")
bic <- mfit_correct(fit_index = "BIC")
aicc <- mfit_correct(fit_index = "AICC")
bicc <- mfit_correct(fit_index = "BICC")

fit_combined <- aic %>%
  select(condition, aic = prop) %>%
  left_join(bic %>% select(condition, bic = prop), by = "condition") %>%
  left_join(aicc %>% select(condition, aicc = prop), by = "condition") %>%
  left_join(bicc %>% select(condition, bicc = prop), by = "condition")

#update table to have condition characteristics#
conditions$condition <- seq(1,36,1)

fit_combined <- fit_combined %>%
  left_join(conditions %>% select(condition, years, csm_dgm, per_var, coh_var),
            by = "condition") %>%
  select(condition, years, csm_dgm, per_var, coh_var, aic, bic, aicc, bicc)


################################################################################
```

```
#combine all results into one workbook and export#

setwd("~/Desktop")

pages <- list("convergence" = converge_wide,
              "rpb_int" = rpb_int,
              "rpb_age" = rpb_age,
              "rpb_age2" = rpb_age2,
              "rpb_female" = rpb_female,
              "rpb_divorce" = rpb_divorce,
              "rpb_divlaw" = rpb_divlaw,
              "rpb_l1res" = rpb_l1res,
              "rpb_perres" = rpb_perres,
              "rpb_cohres" = rpb_cohres,
              "cov_int" = cov_int,
              "cov_age" = cov_age,
              "cov_age2" = cov_age2,
              "cov_female" = cov_female,
              "cov_divorce" = cov_divorce,
              "cov_divlaw" = cov_divlaw,
              "model_fit" = fit_combined)

write.xlsx(pages, file = "Compiled Simulation Results.xlsx",
           numFmt = "0.000")

################################################################################
#visualizations#

#function for dispersion plot#
disp_plot <- function(con, item, labtitle, labs, subtext){
  dat <- fix_eff %>% filter(condition == con & par == item)

  dat$em <- factor(dat$em, levels = paste("cgroup", c(3,5,10), sep = ""),
                   labels = labs)

  intercept <- dat %>%
    group_by(em) %>%
    summarise(intercept = mean(result))

  dat <- dat %>%
    left_join(intercept, by = "em")

  disp <- ggplot(data = dat, aes(x = result, y = model)) +
    geom_point(color = "midnight blue", alpha = 0.5, size = 0.5) +
    facet_wrap(~em, nrow = 3, strip.position = "left") + theme_light() +
    theme(panel.grid.minor = element_blank(),
          panel.grid.major.y = element_blank(),
          axis.ticks.y = element_blank(),
          axis.text.y = element_blank(),
          axis.title.y = element_blank(),
          axis.title.x = element_text(size = 9),
          strip.background = element_rect(fill = "white"),
          strip.text = element_text(color = "black", size = 9),
          strip.text.y = element_text(angle = 180),
          plot.title = element_text(hjust = 0.5, face = "bold", size = 11),
          plot.subtitle = element_text(hjust = 0.5, size = 9)) +
```

```
    geom_vline(aes(xintercept = intercept), color = "red", size = 1) +
    labs(x = labtitle, title = con, subtitle = subtext)

  return(disp)
}

#figure 1#
disp3 <- disp_plot(con = "Condition 3", item = "agez",
         labtitle = expression("Point Estimate of \u03b2"[1]),
         labs = c("3-Year EM\n-0.60", "5-Year EM\n-0.93", "10-Year EM\n-7.69"),
         subtext = "30 Years; 3-Year CSM; Low Period Var.; Low Cohort Var.")

disp21 <- disp_plot(con = "Condition 21", item = "agez",
    labtitle = expression("Point Estimate of \u03b2"[1]),
    labs = c("3-Year EM\n-1.48", "5-Year EM\n-12.44", "10-Year EM\n36.85"),
    subtext = "30 Years; 3-Year CSM in DGM; Low Period Var.; High Cohort Var.")

disp30 <- disp_plot(con = "Condition 30", item = "agez",
    labtitle = expression("Point Estimate of \u03b2"[1]),
    labs = c("3-Year EM\n-2.15", "5-Year EM\n26.55", "10-Year EM\n15.58"),
    subtext = "30 Years; 3-Year CSM in DGM; High Period Var.; High Cohort Var.")

#export plot#
plot_grid(disp3, disp21, disp30,
          nrow = 3,
          align = "v")

dev.new(width = 5, height = 7.5, unit="in", noRStudioGD = TRUE);
last_plot()
ggsave("~/Desktop/Fig1.png",width = dev.size()[1],height = dev.size()[2]);
dev.off()


#figure 3#
disp14 <- disp_plot(con = "Condition 14", item = "divorce",
    labtitle = expression("Point Estimate of \u03b3"[1]),
    labs = c("3-Year EM\n-53.90", "5-Year EM\n2.86", "10-Year EM\n6.41"),
    subtext = "20 Years; 5-Year CSM in DGM; High Period Var.; Low Cohort Var.")

disp22 <- disp_plot(con = "Condition 22", item = "divorce",
    labtitle = expression("Point Estimate of \u03b3"[1]),
    labs = c("3-Year EM\n-59.42", "5-Year EM\n-4.07", "10-Year EM\n-13.15"),
    subtext = "10 Years; 5-Year CSM in DGM; Low Period Var.; High Cohort Var.")

disp33 <- disp_plot(con = "Condition 33", item = "divorce",
    labtitle = expression("Point Estimate of \u03b3"[1]),
    labs = c("3-Year EM\n-63.57", "5-Year EM\n-0.97", "10-Year EM\n-72.85"),
    subtext = "30 Years; 5-Year CSM in DGM; High Period Var.; High Cohort Var.")

#export plot#
plot_grid(disp14, disp22, disp33,
          nrow = 3,
          align = "v")

dev.new(width = 5, height = 7.5, unit="in", noRStudioGD = TRUE);
last_plot()
ggsave("~/Desktop/Fig3.png",width = dev.size()[1],height = dev.size()[2]);
```

```
dev.off()
#figure 6#

#backup original fixed effects results#
fix_eff_backup <- fix_eff

#for purposes of dispersion plot function, rename cohort variance components#
fix_eff$par <- ifelse(fix_eff$par %in% c("cgroup3", "cgroup5", "cgroup10"),
                      "coh_var", fix_eff$par)

#divide variance component by 100,000 for plot scaling purpose#
fix_eff$result <- ifelse(fix_eff$par == "coh_var", fix_eff$result / 100000,
                         fix_eff$result)


disp31 <- disp_plot(con = "Condition 31", item = "coh_var",
  labtitle = expression("Variance Component Estimate, in 100k, of \u03c4"["u"]),
  labs = c("3-Year EM\n-23.43", "5-Year EM\n-0.98", "10-Year EM\n347.99"),
  subtext = "10 Years; 5-Year CSM in DGM; High Period Var.; High Cohort Var.")

disp32 <- disp_plot(con = "Condition 32", item = "coh_var",
  labtitle = expression("Variance Component Estimate, in 100k, of \u03c4"["u"]),
  labs = c("3-Year EM\n-16.29", "5-Year EM\n3.07", "10-Year EM\n447.47"),
  subtext = "20 Years; 5-Year CSM in DGM; High Period Var.; High Cohort Var.")

disp33 <- disp_plot(con = "Condition 33", item = "coh_var",
  labtitle = expression("Variance Component Estimate, in 100k, of \u03c4"["u"]),
  labs = c("3-Year EM\n-19.67", "5-Year EM\n-0.76", "10-Year EM\n531.50"),
  subtext = "30 Years; 5-Year CSM in DGM; High Period Var.; High Cohort Var.")

#export plot#
plot_grid(disp31, disp32, disp33,
          nrow = 3,
          align = "v")

dev.new(width = 5, height = 7.75, unit="in", noRStudioGD = TRUE);
last_plot()
ggsave("~/Desktop/Fig6.png",width = dev.size()[1],height = dev.size()[2]);
dev.off()


#function for zipper plot#
zip_plot <- function(con, item, cov3, cov5, cov10, labtitle, line, subtext){
  temp <- conf_fix %>%
  filter(par == item & condition == con) %>%
  mutate(above_below = ifelse(result <= truth, "below", "above"),
 length = ifelse(above_below == "below" & int_con == 0,
                truth-conf.high, ifelse(above_below == "above" & int_con == 0,
                                        conf.low - truth, conf.high-conf.low)))
  temp <- temp %>%
    group_by(em) %>%
    arrange(desc(int_con), length) %>%
    mutate(sort = row_number())

  temp$em <- factor(temp$em, levels = paste("cgroup", c(3,5,10), sep = ""),
                    labels = paste(c(3,5,10), "-Year EM", sep = ""))
```

```
  lab_text <- data.frame(label = c(cov3, cov5, cov10),
                         em = factor(c("cgroup3", "cgroup5", "cgroup10"),
                              levels = paste("cgroup", c(3,5,10), sep = ""),
                              labels = paste(c(3,5,10), "-Year EM", sep = "")),
                         conf.low = 0, conf.high = 0, int_con = 2)

  zipper <- ggplot(data = temp,aes(x = result, y = sort, xmin = conf.low,
                                   xmax = conf.high, color = factor(int_con))) +
    geom_errorbarh() + facet_wrap(~em) + theme_light() +
    geom_vline(xintercept = line, size = 1.5, color = "grey") +
    theme(panel.grid = element_blank(), legend.position = "none") +
    scale_color_manual(values = c("firebrick", "midnightblue", "black")) +
    theme(axis.text.y = element_blank(), axis.ticks.y = element_blank(),
          strip.background = element_rect(fill = "white"),
          strip.text = element_text(color = "black", size = 9),
          plot.title = element_text(hjust = 0.5, face = "bold", size = 11),
          plot.subtitle = element_text(hjust = 0.5, size = 9),
          axis.title.x = element_text(size = 9),
          panel.border = element_rect(color = "black")) +
    labs(x = labtitle, y = "", title = con, subtitle = subtext) +
    geom_label(data = lab_text,
               mapping = aes(x = -Inf, y = -Inf, label = label),
               hjust = -0.1, vjust = -0.3, size = 3)

  return(zipper)
}


#figure 2#
zip6 <- zip_plot("Condition 6", item = "agez",
     cov3 = "85.15", cov5 = "95.16", cov10 = "40.11",
     labtitle = expression("95% Confidence Interval of \u03b2"[1]),
     line = -5.9251,
     subtext = "30 Years; 5-Year CSM in DGM; Low Period Var.; Low Cohort Var.")

zip24 <- zip_plot("Condition 24", item = "agez",
    cov3 = "76.83", cov5 = "93.50", cov10 = "13.43",
    labtitle = expression("95% Confidence Interval of \u03b2"[1]),
    line = -5.9251,
    subtext = "30 Years; 5-Year CSM in DGM; Low Period Var.; High Cohort Var.")

zip33 <- zip_plot("Condition 33", item = "agez",
    cov3 = "72.06", cov5 = "93.16", cov10 = "11.82",
    labtitle = expression("95% Confidence Interval of \u03b2"[1]),
    line = -5.9251,
    subtext = "30 Years; 5-Year CSM in DGM; High Period Var.; High Cohort Var.")

#export plot#
plot_grid(zip6, zip24, zip33, nrow = 3, align = "v")

dev.new(width = 6.4, height = 8, unit="in", noRStudioGD = T);
last_plot()
ggsave("~/Desktop/Fig2.png",width = dev.size()[1],height = dev.size()[2]);
dev.off()
```

```
#figure 4#
zip13 <- zip_plot("Condition 13", item = "divorce",
    cov3 = "81.21", cov5 = "93.89", cov10 = "88.78",
    labtitle = expression("95% Confidence Interval of \u03b3"[1]),
    line = 3.6598,
    subtext = "10 Years; 5-Year CSM in DGM; High Period Var.; Low Cohort Var.")

zip22 <- zip_plot("Condition 22", item = "divorce",
    cov3 = "89.01", cov5 = "93.87", cov10 = "83.37",
    labtitle = expression("95% Confidence Interval of \u03b3"[1]),
    line = 3.6598,
    subtext = "10 Years; 5-Year CSM in DGM; Low Period Var.; High Cohort Var.")

zip31 <- zip_plot("Condition 31", item = "divorce",
    cov3 = "89.01", cov5 = "93.60", cov10 = "82.26",
    labtitle = expression("95% Confidence Interval of \u03b3"[1]),
    line = 3.6598,
    subtext = "10 Years; 5-Year CSM in DGM; High Period Var.; High Cohort Var.")



#export plot#
plot_grid(zip13, zip22, zip31, nrow = 3, align = "v")

dev.new(width = 6.4, height = 8, unit="in", noRStudioGD = T);
last_plot()
ggsave("~/Desktop/Fig4.png",width = dev.size()[1],height = dev.size()[2]);
dev.off()


#plot for relative bias of period-level predictor#
per_rpb_plot <- rpb_divlaw %>%
  ungroup() %>%
  mutate(Condition = seq(1,36)) %>%
  gather(EM, RPB, cgroup3:cgroup10) %>%
  arrange(Condition) %>%
  mutate(Group = rep(seq(1,12), each = 9))

per_rpb_plot$EM <- factor(per_rpb_plot$EM, levels = paste("cgroup", c(3,5,10),
                                                          sep = ""),
                          labels = paste(c(3,5,10), "-Year EM", sep = ""))



per_rpb_plot <- per_rpb_plot %>%
  mutate(group = ifelse(per_var == "low" & coh_var == "low",
          "Low Period Variability \nLow Cohort Variability",
          ifelse(per_var == "high" & coh_var == "low",
                 "High Period Variability \nLow Cohort Variability",
                 ifelse(per_var == "low" & coh_var == "high",
                        "Low Period Variability \nHigh Cohort Variability",
                        "High Period Variability \nHigh Cohort Variability"))),
         csm_lab = paste(csm_dgm, "CSM in DGM", sep = " "))
```

```
per_rpb_plot$csm_lab<- factor(per_rpb_plot$csm_lab,
                         levels = c("3-Year CSM in DGM",
                                    "5-Year CSM in DGM",
                                    "10-Year CSM in DGM"),
                         labels = paste(c(3,5,10), "-Year CSM in DGM", sep = ""))

#figure 5#
ggplot(data = per_rpb_plot, aes(x = years, y = RPB, group = EM, color = EM)) +
  geom_hline(yintercept = c(-5,5), alpha = 0.8, linetype = "dashed",
             color = "dark grey") +
  geom_line() + geom_point() +
  facet_grid(rows = vars(group), cols = vars(csm_lab)) + theme_bw() +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major = element_blank(),
        legend.position = "bottom",
        legend.background = element_blank(),
        legend.box.background = element_rect(color = "black"),
        strip.text = element_text(size = 10)) +
  labs(x = "Number of Survey Years",
       y = expression("Relative Absolute Bias of \u03b3"[2]),
       color = "Estimating Model") +
  scale_color_manual(values = c("firebrick", "blue4", "darkgoldenrod1")) +
  guides(color = guide_legend(title.position = "top", title.hjust = 0.5))


dev.new(width = 6.4, height = 8, unit="in", noRStudioGD = T);
last_plot()
ggsave("~/Desktop/Fig5.png",width = dev.size()[1],height = dev.size()[2]);
dev.off()
```