4-6-2017

Robustness of the Within- and Between-Series Estimators to Non-Normal Multiple-Baseline Studies: A Monte Carlo Study

Seang-Hwane Joo
University of South Florida, sjoo@mail.usf.edu

Robustness of the Within- and Between-Series Estimators to Non-Normal Multiple-Baseline Studies: A Monte Carlo Study

by

Seang-Hwane Joo

A dissertation submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
in Curriculum and Instruction with an emphasis in
Measurement and Evaluation
Department of Educational and Psychological Studies
College of Education
University of South Florida

Major Professor: John Ferron, Ph.D.
Tasha Beretvas, Ph.D.
Yi-Hsin Chen, Ph.D.
Eun Sook Kim, Ph.D.

Date of Approval:
March 24, 2017

Keywords: Multilevel modeling, Single-case, Non-normality, ReML, MCMC

ACKNOWLEDGEMENTS

I would like to thank my major professor, Dr. John Ferron. I would not have been able to finish this dissertation without his guidance and support. He is a great advisor, researcher, teacher, and mentor. Throughout my doctoral study, he has taught me and inspired me academically and personally. He has guided me on how to conduct research, how to teach courses, and how to treat other people with respect. He inspired me in many ways and it was my greatest honor to have him as my advisor.

I would also like to thank my committee members, Drs. Tasha Beretvas, Yi-Hsin Chen and Eun Sook Kim. Their insightful comments and feedback substantially and significantly improved the quality of my dissertation. They also continuously encouraged and supported me until the very end. Also, I owe special thanks to Dr. Steve Stark. He has challenged me to have a critical perspective and encouraged me to keep moving forward.

I would like to thank my family, my wife Jenny and my son Noel. They have always trusted in me and encouraged me. Also, my gratitude goes to my family in Korea, mother, Sunju Hwang, father, Jongsik Joo and sister, Seungyeon Joo. They have supported me and trusted me unconditionally. Without their love and support, I would not have completed this journey.

Lastly, I thank my heavenly Father for everything He has done for me.

TABLE OF CONTENTS

LISTS OF TABLES

LISTS OF FIGURES

ABSTRACT

In single-case research, multiple-baseline (MB) design is the most widely used design in practical settings. It provides the opportunity to estimate the treatment effect based on not only within-series comparisons of treatment phase to baseline phase observations, but also time-specific between-series comparisons of observations from those that have started treatment to those that are still in the baseline. In MB studies, the average treatment effect and the variation of these effects across multiple participants can be estimated using various statistical modeling methods. Recently, two types of statistical modeling methods were proposed for analyzing MB studies: a) within-series model and b) between-series model. The within-series model is a typical two-level multilevel modeling approach analyzing the measurement occasions within a participant, whereas the between-series model is an alternative modeling approach analyzing participants' measurement occasions at certain time points, where some participants are in the baseline phase and others are in the treatment phase. Parameters of both within- and between-series models are generally estimated with restricted maximum likelihood (ReML) estimation and ReML is developed based on the assumption of normality (Hox, et al., 2010; Raudenbush & Bryk, 2002). However, in practical educational and psychological settings, observed data may not be easily assumed to be normal. Therefore, the purpose of this study is to investigate the robustness of analyzing MB studies with the within- and between-series models when level-1 errors are non-normal. A Monte Carlo study was conducted under the conditions where level-1 errors were generated from non-normal distributions in which skewness and kurtosis of the

distribution were manipulated. Four statistical approaches were considered for comparison based on theoretical and/or empirical rationales. The approaches were defined by the crossing of two analytic decisions: a) whether to use a within- or between-series estimate of effect and b) whether to use REML estimation with Kenward-Roger adjustment for inferences or Bayesian estimation and inference. The accuracy of parameter estimation and statistical power and Type I error were systematically analyzed. The results of the study showed the within- and between-series models are robust to the non-normality of the level-1 error variance. Both within- and between-series models estimated the treatment effect accurately and statistical inferences were acceptable. ReML and Bayesian estimations also showed similar results in the current study. Applications and implications for applied and methodology researchers are discussed based on the findings of the study.

CHAPTER ONE: INTRODUCTION

Single-case research is a type of research for analyzing the effect of an intervention or treatment. Single-case research studies mainly differ from other intervention research studies in that they deal with a single or small number of participants. To evaluate the effect of treatment, repeatedly measured observations of each participant are collected for two distinct phases. Baseline and treatment are two basic phases that comprise the interrupted time series design. The baseline phase is also called the pre-treatment phase and it consists of a series of observations before introducing a treatment or intervention for participants. There are two purposes of the baseline phase in single-case research design: a) observations in the baseline phase provide prior knowledge about each participant's performance and document the need for intervention, and b) observations in the baseline phase establish the basis for which predictions can be made for the participants if the intervention had not been implemented. The treatment phase consists of a series of observations after the introduction of a treatment. Comparing observations between the baseline and treatment phases creates the analysis of the treatment effect in single-case research.

In single-case research, multiple-baseline (MB) is the most widely used design in practical settings (Honor & Odom, 2014). It provides the opportunity to estimate the treatment effect based on not only *within-participant* comparisons of treatment phase to baseline phase observations, but also time-specific *between-participant* comparisons of observations from those that have started treatment to those that are still in the baseline.

Recently, two types of statistical models were proposed for analyzing MB studies: a) within-series (within-participant) model and b) between-series (between-participant) model (Ferron et al., 2014). The within-series model is a typical two-level multilevel modeling approach analyzing the measurement occasions within a participant, whereas the between-series model is an alternative modeling approach analyzing a subset of the participants' measurement occasions, which correspond to certain time points, where some participants are in the baseline phase and others are in the treatment phase. The advantage of the within-series approach is that the treatment effects are estimated with greater precision because all collected observations are used. On the other hand, the between-series approach does not rely on assumptions where time trends are correctly specified. A simulation study found that the within-series model outperformed the between-series model when assumptions are satisfied but the between-series model produced less biased average treatment effects than the within-series model for the conditions where the model is misspecified or event effects are included (Ferron et al., 2014).

Parameters of both within- and between-series models are generally estimated with restricted maximum likelihood (ReML) estimation. ReML is an iterative procedure to find the estimates, which maximize the likelihood function for the variance components of the model (Raudenbush & Bryk, 2002). In general, ReML is developed based on the assumption of normality (Hox, Moerbeek & Van de Schoot, 2010; Raudenbush & Bryk, 2002). However, in practical educational and psychological settings, the normality assumption is not always satisfied (Shadish & Sullivan, 2011; Smith, 2012) and meta-analysis of the single-case research studies has shown that more than 50% of the published single-case research datasets failed the normality test (Parker, 2006; Parker & Vannest, 2009; Parker, Vannest & Brown, 2009; Solomon, 2014). In addition, I conducted a preliminary survey to investigate the normality of the single-case data

published in *Journal of Applied Behavior Analysis (JABA)* from 2014 to 2016. The results from the survey showed that skewness and kurtosis of the MB data ranged from -0.71 to 1.91, and from -1.07 to 3.01, respectively. Details of this survey are discussed in later chapters.

One possible reason for non-normal outcome variables is the scales of measurement. For example, if a scale of measurement is a count or percent, then either positive or negative skewnesss of the data may be observed. If a scale of measurement is binary or categorical, then normality of the data cannot be assumed (Smith, 2012; Shadish, 2014; Shadish, Kyse, & Rindskopf, 2013). The ceiling/floor effects or outliers are another source of creating violation of the normality assumption (Hox et al., 2010; Langford & Lewis, 1998). Outliers could occur due to a momentary or temporary event effects. For example, a momentary event effect occurs if a participant of an intervention study experiences a persoval problem at home at a certain time point and the observed outcome at the next time point is influenced by the event effect either positively or negatively (Ferron et al., 2014).

In principle, analyzing non-normal data with ReML estimation could result in biased estimates of fixed and random effects (Crawford, Garthwaite, Azzalini, Howell & Laws, 2005; Mass & Hox, 2004). Also, standard errors of the parameter estimates tend to be underestimated, resulting in the following consequences; a) the statistical powers to detect fixed and random effects are downgraded, b) Type I error rates for the estimated parameters are inflated and c) inaccurate confidence interval (CI) coverage rates are observed (Crawford et al., 2005; Mass & Hox, 2004; van der Leeden et al., 1997). The same concerns can be applied for the within- and between-series models for MB studies and these problematic consequences of violation of normality could significantly affect the results of the studies.

Therefore, it is important to investigate alternative modeling approaches to handle non-normality of MB data. Also, from a practical perspective, comparing the robustness of multiple modeling methods can provide better insight and practical solutions to applied researchers about how to deal with violation of the normality assumption. For this reason, four statistical modeling approaches are considered for comparison based on theoretical and/or empirical rationales. The approaches are defined by the crossing of two analytic decisions: a) whether to use a within- or between-series estimate of effect and b) whether to use ReML with Kenward-Roger adjustment (Kenward & Roger, 1997) or Bayesian approach (Gelman et al., 2004) for parameter estimation and statistical inferences.

Kenward-Roger adjustment was developed to adjust for small-sample bias. More specifically, the Kenward-Roger method adjusts degrees of freedom to make a better inference for the small-sample size condition. Simulation studies have shown that using Kenward-Roger adjustment has shown a better performance than other degrees of freedom estimation methods for making inferences about the treatment effect in MB studies (Ferron, Bell, Hess, Rendina-Giobioff, & Hibbard, 2009). Also, making inferences with Kenward-Roger adjustment is robust to non-normality in the level-2 errors and misspecification of the level-1 covariance structure in multilevel modeling (Petit-Bois, Baek, Nguyen & Ferron, 2013; Owens & Farmer, 2013).

Bayesian estimation is an alternative way for estimating parameters and making inferences. Over the decades, Bayesian estimation has received increasing attention for estimating the parameters of statistical models, because it is comparably simple to adapt and a better way to deal with more complex models than ReML (e.g., Browne, Draper, Goldstein, Rasbash, 2002; Rindskopf, 2014; Swaminathan, Rogers & Hornor, 2014; Shadish, Rindskopf, Hedges & Sullivan, 2013). Bayesian estimation allows specifying prior distributions to the

parameters in the model. By doing so, researchers can have flexibility to handle parameter estimation under various conditions. For example, it has been shown that Bayesian estimation is beneficial for estimating complex statistical models with small sample sizes due the specification of the prior distribution (Gelman et al., 2014).

Bayesian estimation, previously, has been adapted for the within-series multilevel model and its estimation accuracy was compared with ReML (e.g., Baek, 2015; Moeyaert et al., 2016; Swaminathan, Rogers & Honor, 2014). However, Bayesian estimation has not been employed for the newly proposed between-series model and its estimation efficacy for the model is unknown. It is also unknown whether the between-series model would perform robustly with Bayesian inference compared to Kenward-Roger adjustment under the situations where the normality assumption is violated.

Problem Statement

Although violation of the normality assumption is a potential threat in MB studies, questions about the robustness of statistical models remained unsolved. For example, to what degree would non-normality in level-1 errors be troublesome when MB studies are analyzed with multiple modeling approaches? How much skewness and kurtosis can and cannot be handled using ReML with Kenward-Roger or Bayesian estimation?

The majority of previous applications and methodological studies with MB design assumed that the level-1 errors were normally distributed (e.g., Ferron et al., 2009, Ferron, Farmer, & Owens, 2010, Ferron et al., 2014; Moeyaert, Ugille, Ferron, Beretvas, Van den Noortgate, 2013a, 2013b; Van den Noortgate & Onghena, 2003a, 2003b) and limited research investigated the robustness of the within- and between-series models. For example, the impacts

of level-2 and level-3 error non-normality (Petit-Bois et al., 2013) and level-1 and level-2 error non-normality (Owens & Farmer, 2013) in the within-series model were previously investigated and they found that non-normality does not lead to bias in estimating treatment effects but effect inferences were inaccurate. However, these studies were limited to the within-series model using ReML and they did not compare alternative modeling approaches such as the between-series model or Bayesian estimation. To the best of author's knowledge, no research has compared the performances of various modeling approaches that may be differentially robust to non-normal MB data.

Purpose of Study

The purpose of this study is to investigate the robustness of analyzing MB studies with the within- and between-series models using ReML with Kenward-Roger adjustment or Bayesian estimation when level-1 errors are assumed to be non-normal. A Monte Carlo study was conducted under the conditions where level-1 errors were generated from non-normal distributions manipulating skewness and kurtosis of the residuals' distribution. Fleishman's (1978) power transformation method was used to manipulate skewness and kurtosis of the distribution. To compare various modeling methods, four models are compared: a) two-level within-series model using ReML with Kenward-Roger adjustment (Model 1), b) two-level within-series model using Bayesian method (Model 2), c) between-series model using ReML with Kenward-Roger adjustment (Model 3) and d) between-series model using Bayesian method (Model 4).

The accuracy of parameter estimation and statistical inference were systematically analyzed. Bias, relative bias, root mean square error (RMSE), confidence/credible interval (CI)

coverage rates, CI widths and statistical power/Type I error were examined as a function of specific design factors (number of measurement occasions and participants) and degree of non-normality (amount of skewness, and kurtosis of the distribution). The research questions are described as follows.

Research Questions

1. To what extent are the bias and RMSE for the treatment effect estimates of the within- and between-series models using ReML with Kenward-Roger and Bayesian methods impacted as a function of the skewness and kurtosis?
2. To what extent are the interval estimate coverage rate and width for the treatment effect estimates of the within- and between-series models using ReML with Kenward-Roger and Bayesian methods impacted as a function of the skewness and kurtosis?
3. To what extent are the statistical power and Type I error for the treatment effect estimates of the within- and between-series models using ReML with Kenward-Roger and Bayesian methods impacted as a function of the skewness and kurtosis?
4. To what extent are the bias and RMSE for the parameter estimates other than the treatment effect of the within- and between-series models using ReML with Kenward-Roger and Bayesian methods impacted as a function of the skewness and kurtosis?

Overview of the Study

A Monte Carlo study was conducted to empirically address the issues of violation of the normality assumption in MB studies. Data generation factors included number of measurement occasions and participants, skewness and kurtosis of level-1 error structure, and treatment effect

size. Three levels of number of measurement occasions (10, 20, and 40), two levels of number of participants (4, and 8) four levels of level-1 error skewness (0, 1, 2, and 3) five levels of level-1 error kurtosis (-1, 0, 1, 2, and 4) and two levels of treatment effect sizes (0, and 1) were varied. The analysis factors of the study included four levels of multilevel modeling approaches (Models 1 - 4). Crossing all the data generation factors resulted in a total of 3 (number of measurement occasions) x 2 (number of participants) x 4 (level-1 error skewness) x 5 (level-1 error kurtosis) x 2 (treatment effect) = 240 simulation conditions. Table 1 provides the simulation study conditions. The number of replications was set for 3000 per condition. The dependent variables of the simulation study results were bias, relative bias, RMSE, CI coverage rate, CI width and the statistical power/Type I error of the treatment effect estimate and bias and RMSE of the other parameter estimates including random components in the models.

Significance of the Study

This Monte Carlo study contributes to both applied researchers and methodologists in single-case research. The results of the study can provide applied researchers pragmatic guidance about options when there is violation of the normality assumption, which often occurs in educational and psychological settings. More specifically, the results provide information about analyzing MB studies with multiple statistical modeling approaches and guidelines for applied researchers about how to handle various degrees of violation of the normality assumption.

In addition, this study contributes to methodological literature as well. The between-series model for MB studies is recently proposed and yet, no research has examined the robustness of the between-series model for non-normality of level-1 error. Also, Kenward-Roger adjustment and Bayesian inference methods for the within- and between-series models have not

been compared under non-normal data conditions. The results of the study contribute to methodological literature comparing various modeling approaches for MB studies.

Limitations

The data in this study were simulated based on specific conditions. Those conditions were chosen based on a review of single-case literature. The specific conditions chosen for this study are only some of the possible options. Therefore, the results of this study can only be generalized to studies with similar conditions. Any conclusions beyond the observed conditions should be interpreted with caution.

Table 1.
Simulation Study Design

		Model 1 vs. Model 2 vs. Model 3 vs. Model 4
		Skewness = 0, 1, 2, 3
Participants	Measurement Occasions	Kurtosis = -1, 0, 1, 2, 4
4	10	
	20	
	40	
8	10	
	20	
	40	
4	10	
	20	
	40	
8	10	
	20	
	40	

Definitions of Terms

Bayesian estimation. An estimation method of statistical models where prior information about parameters of the models are taken into account. The estimates of the parameter are computed based on the posterior distribution of the parameters.

Between-series model. A statistical model where the subset of multiple-baseline study is used to compare between participants whose are in the baseline phase to those in the treatment phase.

Bias. A difference between a population parameter value (generating parameter) and an estimated parameter value.

Confidence interval coverage. The proportion of replications in which 95% confidence intervals contain a population parameter value.

Fixed effects. Regression coefficients which present the average effects across level-2 units in multilevel models.

Kenward-Roger. A method that adjusts degrees of freedom of the fixed effects for the small sample size conditions.

Kurtosis. A measure of the heaviness of the tails in a distribution, relative to the normal distribution.

Level-1 error. A residual from the predicted value to the observed value of observations within a level-1 unit in multilevel models.

Level-2 error. A variability across level-2 units in multilevel models.

Multiple-baseline design. A type of single-case research design, which extends the AB design such that the baseline and treatment phases are established for multiple participants, multiple behaviors, or multiple settings.

Multilevel modeling. A statistical model where nested structure data are taken into account for

estimating parameters of the model. It allows researchers to have more than one level of the data structure.

Random effects. The variabilities across level-2 units and level-1 units in multilevel models.

Relative bias. Proportions of bias compared to the population parameter values (generating parameters).

Root mean squared error. The measure of estimation accuracy where squared bias and sampling error are taken into account.

Skewness. A degree of symmetry in a distribution.

Within-series model. Statistical models where multiple-baseline study observations are used to compare those are in the baseline phase to those in the treatment phase.

CHAPTER TWO: LITERATURE REVIEW

The literature review section consists of four parts. First, single-case research design studies are discussed, including an introduction of the single-case research and types of designs, and analysis methods. Second, the within- and between-series models and their application to MB studies are reviewed. Third, the estimation methods and normality assumption are discussed, and finally, alternative Bayesian modeling is described and the relevant literature is reviewed.

Single-Case Research

Single-case research is an intensive study of a case by repeatedly measuring an outcome while altering the conditions under which the case is being observed. In general, the case may be a single participant or a single entity that forms the research group, such as a group of students in a classroom or a family. The outcome variables of a single participant or single entity are then repeatedly measured or quantified over the levels of one or several manipulated independent variables (Onghena, 2005). The independent variables are often manipulated by the researcher, and often, whether participants are observed in intervention (or treatment) or not is the independent variable in single-case research (Kazdin, 2011). The outcome variable (or dependent variable) is the variable participants are measured on (e.g., problem behavior, time on task) and often it is determined by the researchers' theoretical background knowledge or through literature review. The primary purpose of single-case research is, therefore, to investigate the efficacy of

treatment or intervention effect on an outcome variable in which a single participant or entity is involved.

Single-case research studies have been receiving increasing attention recently in educational and psychological studies (e.g., Barlow, Nock, & Hersen, 2009; Ittenbach & Lawhead, 1997; Kazdin, 2011; Kratochwill, 1985; Wacker, Steege, & Berg, 1988). For example, over the last decades, the key terms "single-case" or "single-subject" or "multiple baseline" were exclusively used for citations in the Social Science Citation Index (SSCI), and their applications to educational and psychological studies were substantially increased (Moeyaert et al., 2013a). An increasing number of citations and applications of single-case research occurs because it has contributed greatly to the evidence basis for a variety of practices (Kratochwill & Levin, 2010). Single-case research could be classified as experimental or applied behavior analysis, and it has been applied in various other educational and psychological disciplines. It seeks to establish causal relationships between independent (intervention effect) and dependent variables (outcome measures) with emphasis on understanding individuals' behavior (Kratochwill, 1978; Kratochwill & Levin, 1992). Rather than focusing on the average treatment effect, which is often a primary focus in group comparison experimental design studies, single-case research focuses on case-specific causal effects (Barlow, Nock & Hersen, 2009). In addition, single-case research can be more easily implemented than group experimental design studies for situations where a large number of participants are not available. For example, a researcher may be investigating a low incidence or highly fragmented population, in which a large number of participants may not exist or in which it is expensive to collect a large number of participants for group comparison studies. Also, it is possible that the researcher may be working in an environment (e.g., school or clinical practice) where the logistics and resources are limited. Under these constraints, it is

easier to conduct a study with a small number of participants and examine the intervention effects on a specific case. For these reasons, single-case research studies are sometimes more flexible and feasible than large-group experimental design studies.

Although single-case studies have advantages in applied research settings, they also have their limitations. A natural reaction to small *N* design studies would be questions of generalizability. The interpretation of the results is not easily generalized to the larger population because the study was designed to be case-specific. To address the generalizability issue in single-case studies, researchers have tried to incorporate multiple approaches such as the application of a multiple-baseline design, or replications of the studies. Meta-analysis can be used to assess generalizability of single-case research design studies' results across studies, and to study moderating effects of case and study characteristics.

To examine the causal relationships between independent and dependent variables with single-case research, it is necessary to measure the participant's outcome variables from two distinct phases; a) baseline phase, where outcome variables are measured before the treatment, and b) treatment phase, where outcome variables are measured after the treatment. There are two purposes of the baseline phase in single-case research design: a) observations in the baseline phase provide prior knowledge about the participant's performance and document a problem level of behavior, and b) observations in the baseline phase establish the basis for which predictions can be made for the participant's behavior if the treatment had not been incorporated. The treatment phase consists of a series of observations after the introduction of a treatment. Simply comparing observations between the baseline and treatment phases from a participant often makes the analysis of the effect of a treatment. Figure 1 illustrates typical single-case

observations in baseline phase and treatment phases. They are distinguished by a vertical line with the outcome variable on the y-axis and the time variable on the x-axis.

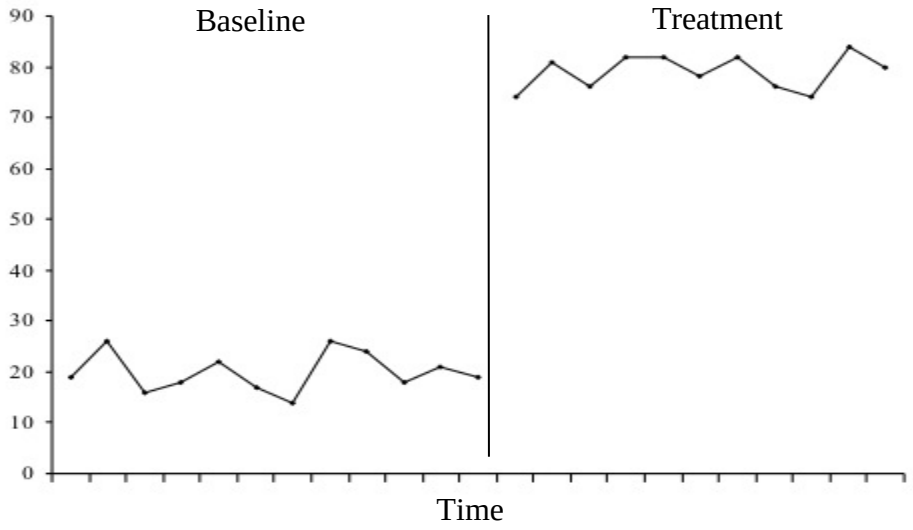


Figure 1. Observations in baseline and treatment phases separated by a vertical line

Types of Single-case Research Designs

In single-case research design studies, several types of single-case research designs were proposed previously, such as an AB design (or interrupted time-series design), a reversal (or withdrawal design; e.g., ABA, ABAB, ABABAB, ABAC designs), and a multiple-baseline design.

AB Design

AB design or interrupted time series (ITS) design is the most basic design in single-case research. AB design consists two phases, phases A and B. Phase A is often referred to as a baseline phase, consisting of a series of observations prior to treatment introduction. Phase B is

referred to as treatment phase, consisting of a series of observations following treatment introduction. Inference about the treatment effect, then, can be made by comparing the difference of outcome variables between the baseline and treatment phases.

AB design is relatively simple and easy to implement because it does not require multiple phases and multiple participants. However, researchers often raise a question about the internal validity of AB design studies because outcome variables could be shifted by something other than true treatment effect (e.g., an event that happened to occur around the time of the treatment; Shadish, Cook, & Campbell, 2002). Suppose a researcher is interested in the effect of a newly implemented learning program on a student's academic performance, and an increase in student academic performance was found after the implementation of the program. Then, it seems natural for the researcher to conclude that the newly implemented learning program is effective in increasing student academic performance. However, one may question if the increment of the student's academic performance may be due to other sources such as academic assistance from a guardian or online learning program that occurs at the same time that the learning program occurs. Therefore, the true effect of the treatment may not be solely observed using AB design.

Reversal Design

To increase the internal validity and conclusions about shifts in time-series data, researchers proposed alternative designs, such as the reversal design. The reversal design is, generally, considered an extension of AB design. One example of the reversal design is ABA design, which increases the phase by withdrawing treatment. Baseline phase (A phase) is additionally included followed by treatment phase (B phase) to observe the pattern of data in which the effect in treatment phase is due to the introduction of treatment. Researchers, then,

expect to observe a similar pattern of observations from the participants as the first baseline phase. To illustrate, Figure 2 represents data using ABA design.

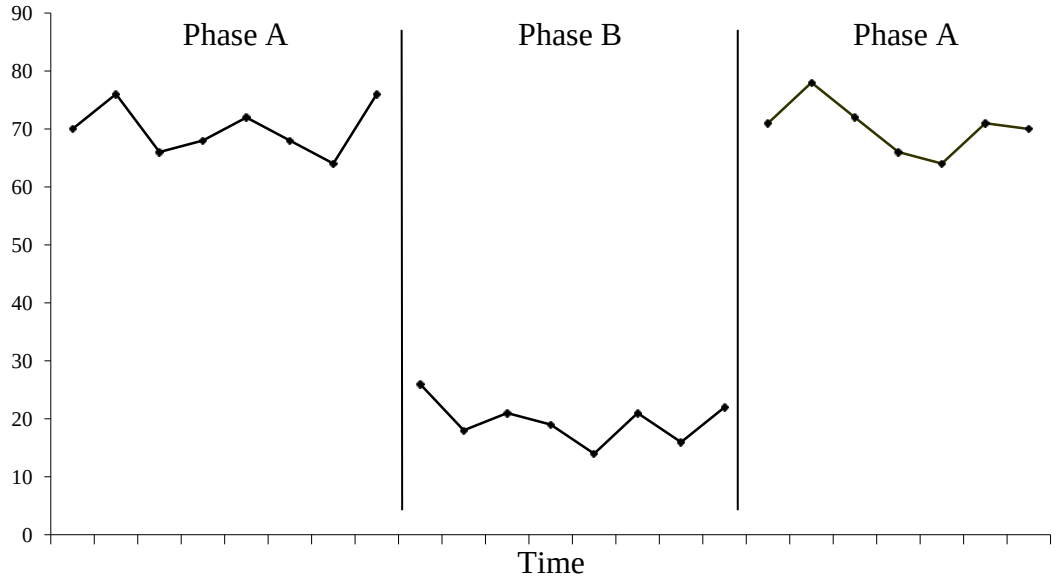


Figure 2. Graphical representation of ABA design with phases separated by a vertical line

By extending more phases through withdrawing and reintroducing treatment, one could have ABAB design. In ABAB design, the second treatment phase is reintroduced to expand the opportunity for researchers to observe the same pattern as the first treatment phase. The advantage of the reversal design is that the reversal design provides the opportunity to clear out event effects in which the shift in outcome variables is caused by something other than the treatment effect (Honor & Odom, 2014).

Although reversal design seems to be an alternative to traditional AB design, there are some practical concerns regarding the reversal design. It is possible that, in some cases, the treatment effect might be permanent or maintained after treatment is withdrawn. It often occurs

in educational settings such as studies involving students' learning. Once students have learned from the treatment phase, it generally is not possible to remove what students learned by withdrawing the treatment.

Multiple-Baseline Design

Another alternative in single-case research studies is multiple-baseline (MB) design. Multiple-baseline is, perhaps, the most widely used design in single-case research (Honor & Odom, 2014; Shadish & Sullivan, 2011). Meta-analysis of single-case research studies reported that 79% of the single-case studies were conducted with some form of multiple baseline design (Shadish & Sullivan, 2011).

MB design is an extension of the AB design that single-case researchers developed to answer the question about internal validity with AB design. MB design extends the AB design such that the baseline and treatment phases are established for either multiple participants, multiple behaviors, or multiple settings. The treatment phases are staggered across time creating different lengths of baseline phases across participants, behaviors, or settings. A graphical representation of a multiple-baseline design with three participants is presented in Figure 3. Similar to the AB design studies, in the MB studies, all participants' observations are repeatedly measured simultaneously to establish the baseline phases for each participant. Once baseline phases are established for all participants, the first participant enters the treatment phase. While the first participant is in the treatment phase, the other participants are still in the baseline phase. A notable change in behavior is expected for the first participant in treatment phase, while the other participants, who are still in baseline are expected to show stable observations in their behaviors. Once sufficient observations are measured for the first participant to evaluate the

effect of the treatment, then the second participant enters the treatment phase while the other participants remain in the baseline phase.

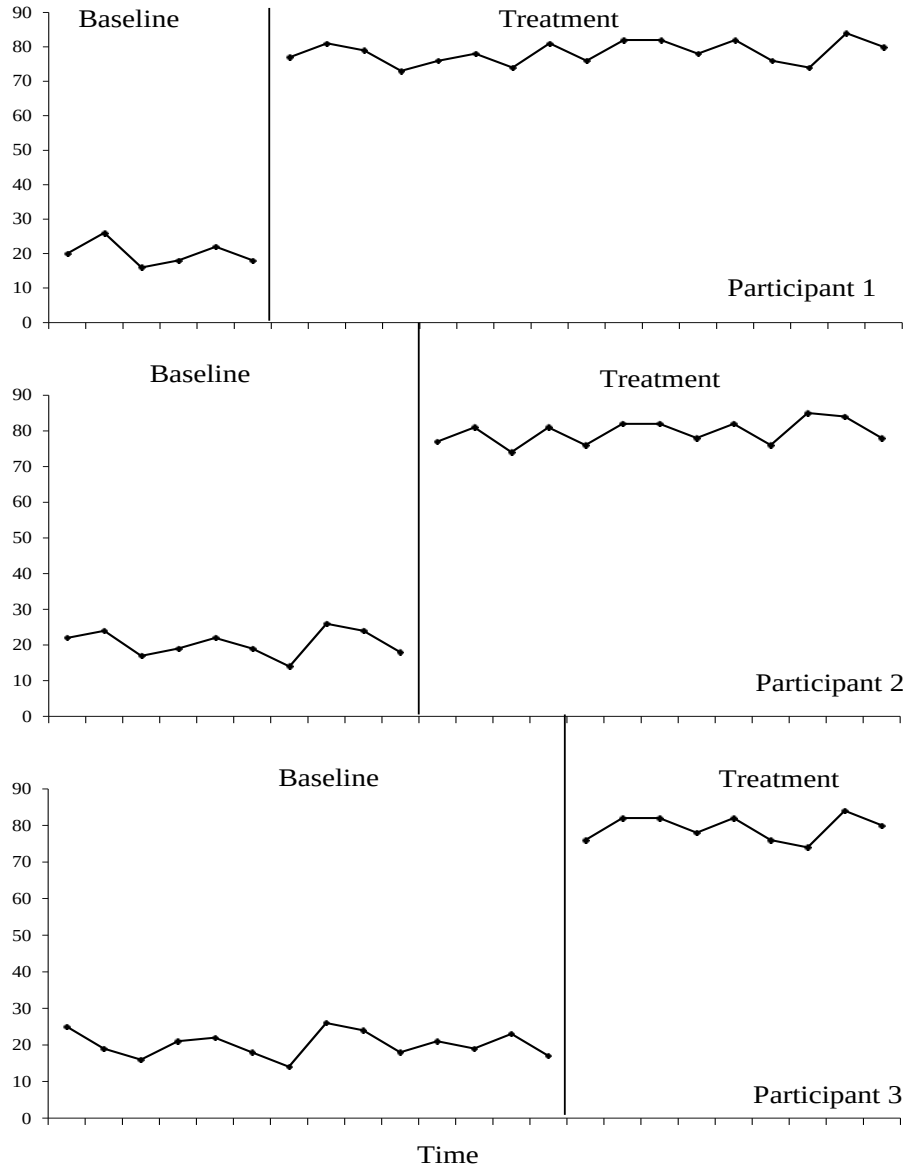


Figure 3. Graphical representation of multiple-baseline design with three participants

Entering the treatment phase at different time points creates the staggered implementation across participants and this makes event effects less plausible, meaning that the change of behavior was more likely due to the treatment effect rather than other event effects such as history or maturation of the participants (Ferron & Scott, 2005). If the change of the participant's behavior was due to a history or maturation effect, then researchers would also expect to observe the change in behavior at the same time for the other participants who are still in the baseline phase. From the analysis perspective, MB design provides the opportunity to analyze the treatment effect and its size of effect based on not only a within-participants (or within-series) comparisons of treatment phase to baseline phase observations, but also a time-specific between-participants (or between-series) comparison of observations from those that have started treatment phase to those that are still in the baseline phase. In addition, MB design also provides the opportunity for researchers to have multiple randomization options. For example, participants can be randomly assigned to a predetermined intervention start time (Wampold & Worsham, 1986) or each participant's intervention start time can be randomly assigned to the participants (Marascuilo & Busk, 1988) or both (Koehler & Levin, 1998). These randomization options increase the internal validity and scientific credibility of the single-case studies (Kratochwill & Levin, 2010).

Analysis Methods of Single-Case Research

Two main streams of analysis methods for single-case research design studies were developed: a) researchers may visually analyze the observations of participants, or b) researchers may analyze the observations using statistical analysis approaches.

Visual Analysis

In visual analysis, researchers visually inspect graphed data focusing on detecting the treatment effect that can be obviously observed through the graphed data (Kazdin, 2011). Visual analysis continues to be the primary method used in the analysis of single-case studies and has been considered as a traditional analysis method because a) it associated with theoretical paradigms of experimental and applied behavior analysis, including in professional journal articles (e.g., *the Journal of Applied Behavior Analysis [JABA]*, or *the Journal of Experimental Analysis of Behavior*), b) multiple and complex factors are taken into account when visual analysts inspect multiple graphs and c) it is an appropriate analysis method for clinical practice, where emphasis on change in the behavior of an individual participant has been the focus (Kratochwill, Hitchcock, Horner, Levine, Odom, Rindskopf & Shadish, 2013; Kratochwill, Levin, Honer & Swoboda, 2014).

Although it is methodologically simple and has a long history with the applied behavior analysis approach, visual analysis includes several limitations regarding scientific evidence. One may have a question about Type I error control and statistical power of visual analysis. Previously, studies have shown that visual analysis inflates the Type I error rates (Fisch, 2001; Matyas & Greenwood, 1990). Alternatively, several methods including training, structured criteria, and masked visual analysis have been suggested and those methods have shown improvement of the accuracy of visual analyses (Ferron, & Jones; 2006; Ferron, Joo, & Levin, 2016; Fisher, Kelley & Lomas; 2003). Another limitation of visual analysis is that a quantified effect size measure of treatment is not provided with which researchers could use to evaluate the treatment. The primary purpose of an intervention study is to examine the effect of the

intervention and how much the intervention affects participants' behavior. Visual analysis may provide whether there is an effect or not but it is limited to the information about how much of an effect occurred. As a result, it is not possible for the results of a visual inspection to be directly connected to quantitative synthesis. When researchers meta-analyze single-case research studies, they rely on effect size measures of the studies (Kratochwill et al., 2014).

Nonparametric Statistics

Nonparametric statistics are, historically, used to evaluate the treatment effect in single-case studies. In general, nonparametric statistical methods do not assume a theoretical distribution (e.g., *t* or F) and as a result, several assumptions in parametric statistics (e.g., normality, equality of variances, and independence of observations) are not required to be assumed (Ferron & Levin, 2014).

The most widely used nonparametric methods in single-case research are permutation and randomization tests. In permutation tests, the test statistics from observed data are first computed and then, the obtained test statistics are compared with an empirical distribution that is formed by either a) all possible permutations of the dataset, or b) computing the test statistics through the sampling with replacement (i.e., resampling method). If the single-case research study included some type of random assignment, and the permutations are based on the possible random assignments, then the permutation test is termed a randomization test (Edgington & Onghena, 2007). Randomization tests are also advantageous because they enhance the internal validity and statistical conclusion validity (Campbell & Stanley, 1963). A strong internal validity leads researchers to conclude with confidence the casual inferences involving interventions and outcomes. In addition, previous methodological research with randomization tests showed that if

randomization tests are conducted with randomization design, Type I errors associated with assessing intervention effect can be well controlled (Ferron, Foster-Johnson & Kromrey, 2003).

However, some concerns with permutation and randomization tests still exist, including sensitivity of statistical power, heavy computational demands and feasibility of randomization (Ferron & Levin, 2014). Previous studies showed that the statistical power of single-case randomization tests varies across multiple factors, from unacceptably low to reasonably high. These multiple factors include the types of single-case designs, randomization methods, amount of data collected, degrees of autocorrelation, and effect sizes (e.g., Ferron & Onghena, 1996; Ferron & Sentovich, 2002; Ferron & Ware, 1995). Also note that although randomization tests provide a *p*-value for significance testing, the effect size measures are not estimable from the randomization tests. The complexity of the randomization procedure is another drawback of these nonparametric tests. Although some user-friendly applications for a standard permutation test are available, most applications of true randomization tests to single-case studies require specialized software or programming scripts. In addition, when the number of observations in single-case data is large, randomization procedures may require heavy computations (Ferron & Levin, 2014).

Non-overlap Statistics

Non-overlap statistics are also heavily-used statistical methods which can provide the numerical or quantified size of a treatment effect. Non-overlap statistics were developed based on the percent of non-overlapping data (Mastropieri & Scruggs, 1985). Various types of non-overlap statistics were proposed and developed previously by a number of researchers in single-case studies: a) percent non-overlapping data (PND; Scruggs, Mastropieri & Casto, 1987), b)

percent exceeding median data (PEM; Ma, 2006), c) percent zero data (PZD; Scotti, Evans, Meyer, & Walker, 1991), d) percentage of all non-overlapping data (PAND; Parker, Hagan-Burke & Vannest, 2007), e) non-overlap of all pairs (NAP; Parker & Vannest, 2009) and f) Tau-U (Parker, Vannest, Davis & Sauber, 2011).

Non-overlap statistics are appealing because a) they were developed without any assumptions about the distribution of the data, so that they could provide robust statistics to handle typical single-case data, and b) they are directly interpretable and appear to be accessible enough for single-case researchers to use (Parker, Vannest & Davis, 2014). However, non-overlapping statistics also include several limitations including ceiling effect, sensitivity to outliers and the assumption of no trend in baseline and treatment (Parker & Vannest, 2009).

Single-level Regression Analysis

One of the parametric statistical analysis methods for single-case data is single-level regression model (e.g., Gottman & Glass, 1978; Huitema & McKean, 2000; Maggin, Swaminathan, Rogers, O'Keeffe, Sugai, & Horner, 2011). To estimate the treatment effect, each participant's observation is modeled by a dichotomized variable *phase* (e.g., *phase* = 0 if observation is from baseline phase and *phase* = 1 if observation is from treatment phase) and the estimated coefficient of the *phase* variable would provide the size of the treatment effect in the model. Equation 1 represents the most basic form of the single-level regression model.

$$Y_i = \beta_0 + \beta_1 * phase + e_i\ , \quad e_i \sim N(0, \sigma^2), \tag{1}$$

where Y_i is participant's observations at i^{th} time point, β_0 is an average observations in baseline phase and β_1 is an estimated shift in level from baseline to treatment phases (i.e., treatment effect). Also, ε_i indicates the error term for the i^{th} time point observation and they are assumed to

be normally distributed with zero mean and variance σ^2. Note that it is possible to model time trend variable for the trend effect in baseline phase or treatment phase or both. Equation 2 represents the single-level regression model for the treatment effect, time trend effect in both baseline and treatment phases.

$$Y_i = \beta_0 + \beta_1 * phase + \beta_2 * time + \beta_3 * (phase * time) + e_i \ , \ e_i \sim N(0, \sigma^2) \qquad (2)$$

where, β_0 is an average observations in baseline phase, β_1 is a treatment effect (i.e., the expected difference between behavior in the treatment and baseline condition at time = 0), β_2 is a trend effect in baseline phase, and β_3 is a trend effect (i.e., the difference in trends between treatment and baseline phases). In general, assumptions of the single-level regression model include independence of observations, homogeneity of variances between baseline and treatment phases, and error terms are normally distributed with zero mean and variance σ^2. However, previous studies examined and developed the more complex single-level regression models and indices regarding the assumptions of variance. For example, generalized least square (GLS) regression model was adapted to estimate the dependent and heterogeneous structures of residual variances (Maggin, et al., 2011), and modified R-square indices were introduced to address the autocorrelation structure in single-case data (Beretvas & Chung, 2008).

Single-level regression analysis is an adequate method because it has flexibility to model not only linear time trend of the treatment effect but also quadratic growth curve or non-linear growth curve. Another significant advantage of single-level regression analysis is that it allows researchers to make an inference about the treatment and time trend effects. Although other statistical analysis methods (e.g., non-overlap statistics) of single-case studies have several advantages for estimating the treatment effect (e.g., relatively straightforward computation and interpretation), they are limited if researchers desire to obtain statistical inferences including

interval estimates or statistical significance tests because they are developed based on a series of assumptions (e.g., independence and homogenous variances). Single-level regression analysis, however, provides greater flexibility for researchers, in that they can estimate not only the size of the effect but also test hypotheses and create confidence intervals under a variety of different assumptions (e.g., independent versus autocorrelated, homogeneous versus heterogeneous).

Single-level regression is a well-suited analysis method for studies with AB or reversal designs involving one participant. However, when multiple participants are involved in single-case studies, such as MB design, replicated ABAB, or replicated alternating treatment design studies, a single-level regression model for each participant might not be optimal if researchers are interested in variability across participants as well as average treatment effects for the study. To analyze the multiple-participant single-case studies, multilevel modeling has been suggested as an analysis method (Rindskopf & Ferron, 2014; Shadish & Rindskopf, 2007; Van den Noortgate & Onghena, 2003a, 2003b, 2007, 2008).

Multilevel Modeling

In educational and psychological research settings, multilevel modeling became popular because it takes a nested structure of data into account. Multilevel models are specifically developed for analyzing hierarchical structure data where lower-level units are nested in higher-level units. These hierarchical structure data are often found in behavioral and social science studies. For example, in educational settings, students are nested in teachers and teachers are nested in schools. Similarly, MB studies can be considered as hierarchical structure such that observations are repeatedly measured over time within a participant and multiple participants are involved in a study.

Analyzing MB studies with multilevel modeling includes various advantages over the multiple single-level modeling approach. For example, multilevel modeling provides not only individual participants' treatment effect estimates, but also the average treatment effect estimate across participants. Multilevel modeling further provides the inference about the average treatment effect estimate as well. Another advantage of using multilevel modeling for MB studies is that researchers could obtain the average variance estimates for within individual participants' observations as well as between participants' observations.

Using multilevel modeling for MB studies also includes limitations. For example, accurate parameter estimates and inferences can be obtained when several assumptions are satisfied. The assumptions include homogeneous variance across two phases (baseline and treatment), homogeneous variance across multiple participants, normality of observations, and correct specification of the fixed effect parameters and random components in the model.

Within-Series Model

The within-series model is a type of multilevel models, which estimates the treatment effect using series of within-participants' observations in MB studies. In typical MB studies, multiple observations are measured within a participant and multiple participants are included in the study. This hierarchical structure can be analyzed with two-level multilevel modeling. That is, participants' observations or measurement occasions are considered as first-level units, and participants are considered as second-level units. To illustrate the variability within and between participants, a simple form of multilevel model, where a treatment effect is used as a predictor in level-1, is described in Equations 3.1-3.3. Equation 3.1 describes participants' measurement occasions for the baseline and treatment phases, respectively, along with the average variation

within participants. Equations 3.2 and 3.3 describe the variation across participants for the predictors in Equation 3.1. Similar to the single-level regression model, a predictor *Phase* is included as a dichotomous variable separating baseline (*Phase* = 0) and treatment (*Phase* = 1) phases. The equation for level-1 is described as follows:

$$Y_{ij} = \beta_{0j} + \beta_{1j} Phase_{ij} + e_{ij}\ , \qquad e_{ij} \sim N(0, \sigma^2) \tag{3.1}$$

Y_{ij} is i^{th} time point measurement occasion for j^{th} participant and β_{0j} and β_{1j} are the intercept and slope effects in the model. Note that coefficients β_{0j} and β_{1j} are allowed to vary across participants and will be described in level-2 equations. Because the *Phase* variable separates the baseline and treatment phases, baseline observations are modeled by β_{0j} with random error e_{ij} and treatment observations modeled by β_{0j} plus β_{1j} with random error e_{ij}. That is, the observations in the treatment phase are expected be either higher or lower than those in the baseline phase by β_{1j} and it is considered as the treatment effect for j^{th} participant. The random error e_{ij} is generally assumed to be independent for each observation and homogeneous across phases. It is also assumed that random errors, e_{ij} are normally distributed with variance, σ^2.

For the level-2 equations, the variation of both coefficients β_{0j} and β_{1j} across participants is described as follows:

$$\beta_{0j} = \gamma_{00} + u_{0j} \tag{3.2}$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \tag{3.3}$$

γ_{00} and γ_{10} are intercept and u_{0j} and u_{1j} are error terms for the model, respectively. From Equations 3.2 and 3.3, baseline observations β_{0j} for j^{th} participant is divided into expected value γ_{00} plus its randomness u_{0j}. Similarly, the treatment effect β_{0j} for j^{th} participant is translated into some constant γ_{10} plus its randomness u_{1j}. Note that γ_{00} is the average baseline observation

across participants and $\gamma_{00} + \gamma_{10}$ is the average treatment observation across participants, so that γ_{10} is the average amount of shift in level from baseline to treatment phases across participants (i.e., treatment effect). The randomness u_{0j} represents the amount of variation between participants for the average baseline observation and u_{1j} represents the amount of variation between participants for the treatment effect. Error structure in level-2 is expressed in matrix form because error terms for intercept and slope are allowed to be correlated. That is, u_{0j} and u_{1j} are distributed as multivariate normal distribution with zero-vector mean and 2 by 2 variance-covariance matrix $\boldsymbol{\Sigma_u}$.

$$\boldsymbol{u} \sim MVN(\mathbf{0}, \boldsymbol{\Sigma_u})$$

where, $\boldsymbol{\Sigma_u} = \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}$ with $\tau_{01} = \tau_{10}$. τ_{00} is variance across participants in the baseline means, τ_{11} is treatment effect variance across participants, and τ_{10} is covariance between baseline level variation and treatment effect variation across participants. It is also possible to constrain the covariance parameters τ_{01} and τ_{10} equal to zero, in other words, $\boldsymbol{\Sigma_u} = diag(\tau_{00}, \tau_{11})$. Note that this constraint indicates that the baseline intercept and treatment effects are assumed to be independent for each participant. Combining level-1 and 2 equations results in the final multilevel model described as follows.

$$Y_{ij} = \gamma_{00} + \gamma_{10} Phase_{ij} + u_{0j} + u_{1j} Phase_{ij} + e_{ij}, \quad e_{ij} \sim N(0, \sigma^2) \qquad (4)$$

Note that γ_{00} and γ_{10} are called fixed effects and τ_{00}, τ_{01}, τ_{11}, and σ^2 are called random components of the model.

Parameters in multilevel modeling can be extended depending on model predictors. Similar to a single-level regression model, it is possible to include additional predictor variables at different levels of the model. To evaluate the trend effects in both baseline and treatment phases, a time variable (*Time*) and its product with the treatment variable (*Time*Phase*) can be

included as predictors in the level-1 equation. Regardless of how time is measured, the *Time* variable is often centered for each participant where *Time* = 0 at the time at which the researcher wants to estimate the treatment effect. The extended equation is described as follows:

$$Y_{ij} = \beta_{0j} + \beta_{1j}Phase_{ij} + \beta_{2j}Time_{ij} + \beta_{3j}(Time_{ij} * Phase_{ij}) + e_{ij},\ e_{ij} \sim N(0, \sigma^2) \quad (5)$$

where, β_{0j} is intercept, β_{1j} is the time specific treatment effect, β_{2j} is the trend in the baseline phase and β_{3j} is the change in trend between the treatment and baseline phase for the j^{th} participant. Since coefficients β_{0j}, β_{1j}, β_{2j} and β_{3j} are random across participants and allowed to be correlated, further equations can be expressed as follows.

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (6.1)$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (6.2)$$

$$\beta_{2j} = \gamma_{20} + u_{2j} \quad (6.3)$$

$$\beta_{3j} = \gamma_{30} + u_{3j} \quad (6.4)$$

where, γ_{00}, γ_{10}, γ_{20}, γ_{30} are average values for β_{0j}, β_{1j}, β_{2j} and β_{3j}, and u_{0j}, u_{1j}, u_{2j} and u_{3j} are error terms, respectively. Similar to the previous multilevel model, γ_{00} is average baseline observation, γ_{10} is average treatment effect, γ_{20} is average trend in baseline and γ_{30} is the across participant average change in trend between treatment and baseline phases. Furthermore, u_{0j}, u_{1j}, u_{2j} and u_{3j} are variations for each effect across participants. They are assumed to be multivariate normal distribution with zero-vector mean and 4 by 4 variance-covariance matrix $\boldsymbol{\Sigma_u}$ where,

$$\boldsymbol{\Sigma_u} = \begin{pmatrix} \tau_{00} & \tau_{01} & \tau_{02} & \tau_{03} \\ \tau_{10} & \tau_{11} & \tau_{12} & \tau_{13} \\ \tau_{20} & \tau_{21} & \tau_{22} & \tau_{23} \\ \tau_{30} & \tau_{31} & \tau_{32} & \tau_{33} \end{pmatrix}$$

Note that $\tau_{ij} = \tau_{ji}$ for $i \neq j$ and τ_{00}, τ_{11}, τ_{22} and τ_{33} are variances for baseline observations, treatment effects baseline trends, and trend changes, respectively. Off-diagonal elements of $\boldsymbol{\Sigma_u}$ are covariance between effects described above. If fixed effects of the model are assumed to be independent then, variance-covariance matrix becomes $\boldsymbol{\Sigma_u} = diag(\tau_{00}, \tau_{11}, \tau_{22}, \tau_{33})$. Combining level-1 and 2 equations results the final multilevel model as described as follows.

$$Y_{ij} = \gamma_{00} + \gamma_{10}Phase_{ij} + \gamma_{20}Time_{ij} + \gamma_{30}\left(Time_{ij} * Phase_{ij}\right)$$
$$+u_{0j} + u_{1j}Phase_{ij} + u_{2j}Time_{ij} + u_{3j}\left(Time_{ij} * Phase_{ij}\right) + e_{ij}, \quad e_{ij} \sim N(0, \sigma^2) \quad (7)$$

Note that γ_{00}, γ_{10}, γ_{20} and γ_{30} are fixed effects and τ_{ij} and σ^2 are random components of the model.

Van den Noortgate and Onghena (2003a, 2003b) demonstrated the use of a two-level model as described above, and noted that if data from several of such single-case studies are combined, a three-level model is recommended to model variability in scores at each of three levels: scores may vary over measurement occasions within participants (level-1), between participants from the same study (level-2) and between studies (level-3). This meta-analytic three-level modeling approach was, recently, investigated to examine the efficacy of the use of the multilevel modeling approach for synthesizing single-case data (Moeyaert, et al., 2013a, 2013b, 2014).

With within-series multilevel modeling, it is possible to include dependent level-1 error structures in the model. A variety of alternative error structures have been suggested for handling the level-1 error dependency in MB studies (Baek & Ferron, 2013; Petit-Bois, 2014) including first-order autoregressive models, AR(1), higher-order autoregressive models, first-order moving average models, MA(1), and autoregressive moving average (ARMA) models. Simulation studies have found that estimation of the treatment effect is unbiased and inferences are

relatively robust to misspecification of the correlation structure of the level-1 residuals (Petit-Bois, 2014; Petit-Bois, Baek, Van den Noortgate, Beretvas, & Ferron, 2016). In addition, it is also possible to include a heterogeneous level-1 error structure in the model. Past simulation studies also supported the viability of estimating separate variances for the baseline and treatment phases when the treatment alters the variance (Bunuan, Hembry, & Beretvas, 2013; Joo & Ferron, 2016). Also, modeling heterogeneous variance across participants has been investigated (Baek & Ferron, 2013) for the within-series model.

Between-Series Model

Ferron et al. (2014) proposed the between-series model as an alternative modeling approach to estimate the treatment effect in MB studies. To illustrate the data used in the between-series model, Figure 4 provides a graphical representation of the simulated data with the four participants in the between-series model. As Figure 4 shows, the observations in the enclosed vertical boxes are used to make comparison between baseline and treatment phases. Dichotomous variable D_{ij} can be created such that $D_{ij} = 0$, if the i^{th} observation for j^{th} participant is not in the enclosed box and $D_{ij} = 1$, otherwise. This allows separating observations that are used for the within-series model and observations that are used for the between-series model. The subset of observations for the between-series model is purposely selected based on the time points where one observation is undergoing the treatment phase whereas the other observations are still in the baseline phase. In Figure 4, there are four different time points that participants are entering the treatment phases and to make comparison of baseline and treatment phases, the observations following the first 3 time points were used (see the three pairs of enclosed boxes in Figure 4). Also note that, from the second pair of vertical boxes, the observations have been

more than 3 time points after the treatment starts is not included. This is because those observations may include the time trend effect as well as immediate level effect (Ferron et al. 2014).

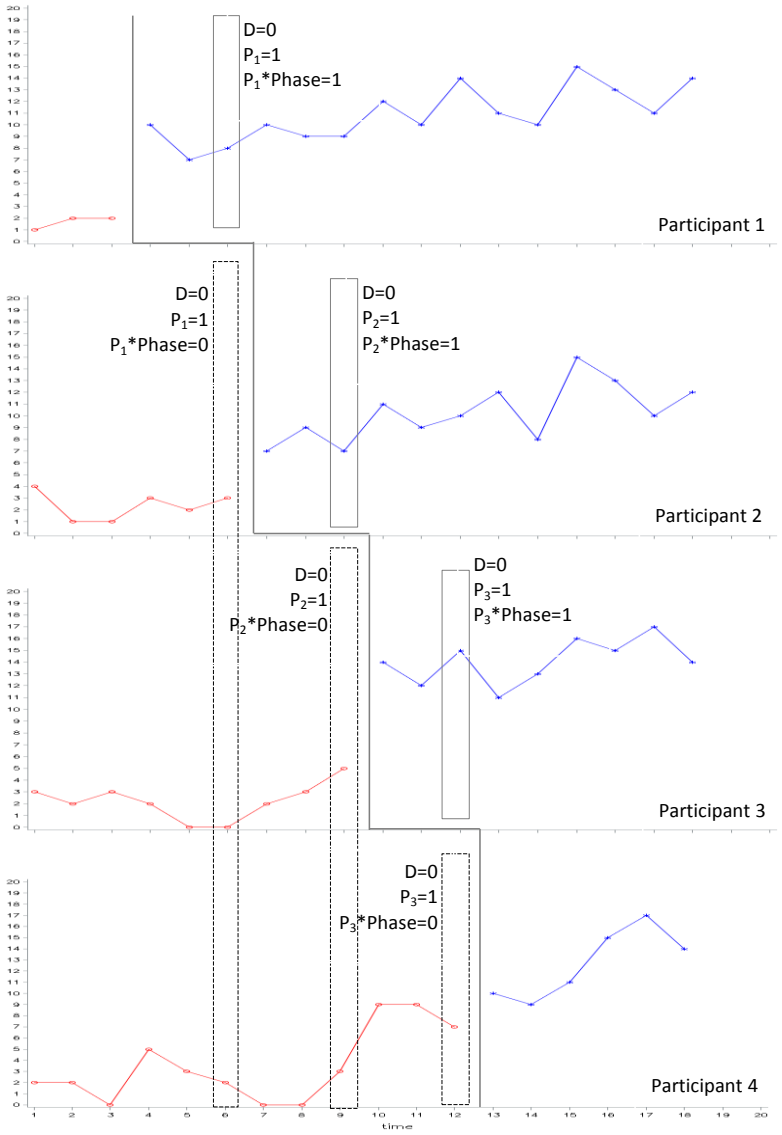


Figure 4. Graphical representation of the data points for the between-series model

The between-series model also uses a series of dichotomously coded variables (P_{1ij}, P_{2ij}, … P_{kij}). P_{kij} is coded 1 if an i^{th} observation for j^{th} participant is at the k^{th} vertical enclosed box and P_k = 0, otherwise. Similar to the within-series model, *Phase* variable is also specified as dummy variable in the between-series model (e.g., 0 = baseline, 1 = treatment) so that interaction term ($P_k * Phase$) can separate observations for the baseline phase, and observations for the treatment phase. In sum, the between-series model equation can be expressed as follows.

$$Y_{ij} = \sum_{k=1}^{K}\left(\beta_k P_{kij} + \beta_{K+k} P_{kij} Phase_{ij}\right) + e_{ij} \quad (8)$$

Note that coefficient β_{K+1} is the treatment effect estimate after the first participant entered the treatment phase, while the other participants still stay in the baseline phase. Similarly, β_{K+2} is the between-series treatment effect estimate after the second participant entered the treatment phase, while the other participants are in the baseline phase. The between-series model in Equation 8 individually estimates the treatment effect for each k^{th} time point. To estimate a single quantity of the treatment effect averaging across the *k* time points (i.e., pooled estimate), Equation 8 can be modified to have a common between-series effect.

$$Y_{ij} = \sum_{k=1}^{K}\left(\beta_k P_{kij}\right) + \beta_{K+1} P_{kij} Phase_{ij} + e_{ij}, \quad (9)$$

where β_{K+1} is the pooled treatment effect estimate across participants. Because of the possibility of time trend effect, baseline observations for each k^{th} time point are still separately estimated. The error term, e_{ij}, in the between-series model represents the combined variabilities for the observations within a participant and between participants. Similar to the within-series model, it is also possible to model various error structures. If researchers assume homogeneous variance across baseline and treatment phases and across k^{th} time points, then single variance would be modeled [i.e., Var(e_{ij}) = σ^2]. If researchers assume heterogeneous variances across two phases, then distinct variances would be modeled [i.e., Var($e_{ij(m)}$) = σ_m^2, where *m* represents phases] and

if researchers assume heterogeneous variances across k^{th} time points, then multiple variances would be modeled [i.e., $\mathrm{Var}(e_{ij(k)}) = \sigma_k^2$, where k represents k^{th} time point for each enclosed box].

The motivations behind the development of the between-series model are twofold: a) it may not be easily assumed that the time trend in the model is of a specific form (e.g., linear) and b) to avoid bias of the treatment effect which can be caused by effects of factors other than the treatment (Ferron et al., 2014). Previously, to model non-linear trajectories in outcome variables, researchers examined various functional forms of within-series models to fit the data best. For example, when outcome variables reach an asymptotic line at the end of the treatment phase, then the growth in the treatment phase can be modeled as a logistic function (Hembry, Bunuan, Beretvas, Ferron & Van den Noortgate, 2014). When the outcome is a count or a rate, then it was suggested that the log function would be appropriate for a non-linear trajectory in multilevel modeling (Shadish et al., 2013; Shadish & Rindskopf, 2007; Shadish, Rindskopf & Hedges, 2008). However, these alternative modeling approaches make sense only when researchers correctly specify the model. Unfortunately, in practical research settings, researchers rarely know the correct model with any confidence. The between-series model can help solve this problem without assuming any functional forms of the model. In addition, when series of observations are analyzed within a participant (i.e., within-series model), the estimated treatment effect can be biased if an event effect is present. The between-series model, however, provides stronger evidence that the changes of participants' observations are due to the treatment effect not to some event effect other than the treatment, by comparing those participants in treatment to those still in baseline (Ferron et al., 2014). A simulation study found that the within-series model outperformed the between-series model when assumptions are satisfied but the between-series

model produced less biased average treatment effects than the within-series model for the conditions where the model is misspecified or event effects are included (Ferron et al., 2014).

ReML Estimation and Inference

Parameters of the within- and between-series models are generally estimated with either maximum likelihood (ML) or restricted maximum likelihood (ReML) estimation. ML estimates both fixed effects and random components of the model simultaneously whereas ReML estimates fixed effects while restricting random components of the model during the estimation procedure. Both ML and ReML are iterative procedures to find the estimates, which maximize the likelihood function of the model (Raudenbush & Bryk, 2002). Previous research that investigated the efficacy of ML and ReML reported that ReML produced more accurate fixed effect estimates than ML and with as few as 6 level-2 units, and the variance components of ReML were estimated with better precision than ML (Browne & Draper, 2000). Also, Browne and Draper (2000) further noted that as the number of level-2 samples increase, both ML and ReML produce reasonable variance component estimates. With regards to statistical inference of the estimates, it was found that standard errors are more accurately estimated with ReML than ML, but still a sufficient number of level-2 units is required to obtain reliable inferences (Maas & Hox, 2004; van der Leeden & Busing, 1994).

Small sample size, especially for level-2 units, is problematic because both ML and ReML were developed based on large-sample theory. As previous research has shown, the recommended sample size for the level-2 unit is at least 30 to obtain unbiased parameter estimates and reliable inferences (Hox, 1998, Maas & Hox, 2004). More specifically, when the level-2 sample size is relatively small (less than 30), the fixed effects of the model are estimated

with no biases, but the variance component estimates are severely biased (Raudenbush & Bryk, 2002). Previous simulation studies also support that substantial bias in the estimates of the variance component when level-2 sample size is less than 30 (Bell, Morgan, Schoeneberger, Kromrey, & Ferron, 2012). Furthermore, statistical inferences for the fixed effects, including the confidence interval and statistical power, would be inaccurate because standard errors of the estimates are underestimated (Mass & Hox, 2004).

In general, it is not common for MB studies to have a large number of participants (i.e., level-2 unit). In fact, very few research studies would involve 30 or more participants in a single MB study (e.g., Koutsoftas, Harmon & Gray, 2009). Thus, to obtain more reliable statistical inferences for the treatment effect estimate using multilevel modeling, it is necessary to incorporate a small sample size adjustment.

Kenward-Roger Degrees of Freedom Adjustment

In the past multilevel modeling with MB studies, various small sample size adjustment methods were introduced to obtain more reliable statistical inferences for the fixed effect estimates. For example, Ferron et al. (2009) compared several small sample size adjustment methods for multilevel modeling in the context of MB studies. They considered five adjustment methods; containment, residual, between-within, Satterthwaite (Satterthwaite, 1946), and Kenward-Roger (Kenward & Roger, 1997). These small sample size adjustments are distinguished from one another in terms of how they compute degrees of freedom for the fixed effects. Relatively simple methods, residual and containment, compute the degrees of freedom as follows:

$$df_{containment} = n_2(n_1 - p) \quad (10)$$

$$df_{residual} = n_2 n_1 - 1 \quad (11)$$

where n_2 is the number of level-2 unit, n_1 is the number of level-1 unit and p is the number of fixed effect parameters in the model. The between-within degrees of freedom method, in general, partitions the residual degrees of freedom into between- and within-participant portions. However, in multilevel modeling with MB studies, all of the residual degrees of freedom are given to the within-participants so that the between-within method is essentially the same as the residual method (Ferron et al., 2009, 2010). These simple computations for the degrees of freedom tend to be overestimated when a more complex variance structure is used, thereby inferences about the estimates would be unreliable. On the contrary, the Satterthwaite and Kenward-Roger methods approximate the degrees of freedom in accordance with the complex variance structure of the observed data. The degrees of freedom approximation using the Satterthwaite method is given as follows.

$$df_{Satterthwaite} = \frac{2\left(c'\widehat{\Sigma_{\hat{\beta}}}c\right)^2}{\left[var\left(c'\widehat{\Sigma_{\hat{\beta}}}c\right)\right]}, \quad (12)$$

where c is defined as $c'\beta = 0$ and $\widehat{\Sigma_{\hat{\beta}}}$ is estimated variance-covariance matrix of $\hat{\beta}$, defined as $\widehat{\Sigma_{\hat{\beta}}} = \left(X\hat{V}^{-1}X'\right)^{-1}$, where $\hat{V}^{-1}$ is inverse variance-covariance matrix and X is design matrix of the fixed effects. The Kenward-Roger method is an extension of the Satterthwaite method. In the Satterthwaite method, the degrees of freedom approximation is adjusted for a small sample size bias (Ferron et al., 2009, 2010). The small sample bias adjustment is made by replacing $\widehat{\Sigma_{\hat{\beta}}}$ with adjusted $\widehat{\Sigma_{\hat{\beta}}^*}$, where $\widehat{\Sigma_{\hat{\beta}}^*}$ is a bias-adjusted estimator of the precision of $\hat{\beta}$ (Kenward & Roger, 1997).

Ferron et al. (2009) showed that the Kenward-Roger and Satterwaite methods for estimating degrees of freedom are preferable to any other method when the within-series

multilevel model was used for MB studies. Although a relatively small number of participants was used, the Kenward-Roger method produced the unbiased treatment effect estimates and their confidence intervals were close to the nominal level (alpha = .05). Mean square error estimates for the treatment effect were also reasonable across conditions as well. In addition, Bell et al., (2012) also found that unbiased fixed effect estimates and accurate Type I error rates using the Kenward-Roger method when sample size was as low as 10.

However, note that variance components in multilevel modeling are still problematic with small sample size because the Kenward-Roger method does not apply for the variance component estimates. Substantial biases in the variance component estimates are consistently found in previous research with small sample size (Bell et al., 2012; Clarke & Wheaton, 2007; Ferron et al., 2009; Moeyaert et al., 2013a, 2013b).

Violation of Normality Assumption for MB Studies

In addition to the estimation being based on large sample size or asymptotic theory, a critical assumption associated with multilevel modeling estimation, is normality of the data (Hox et al., 2010; Raudenbush & Bryk, 2002). Normality is generally defined as whether the theoretical distribution where residuals are assumed to be drawn from is a "bell-shaped" curve (Cohen, Cohen, West, & Aiken, 2003). The levels of normality can be measured by two well-known moment statistics: skewness and kurtosis. Skewness indicates a lack of symmetry in a distribution. Data from a right-skewed (skewed to the right) distribution have values that are bunched together below the mean, but have a long tail above the mean. Similarly, data from a negatively skewed (skewed to the left) distribution have values that are bunched together above the mean, but have a long tail below the mean. On the other hand, kurtosis is a measure of the

heaviness of the tails in a distribution, relative to the normal distribution. A distribution with positive kurtosis (leptokurtic) is light-tailed relative to the normal distribution, while a distribution with negative kurtosis (platykurtic) is heavy-tailed relative to the normal distribution.

This normality assumption is critical for multilevel modeling because both ML and ReML are developed in accordance with the assumption of normality (Eliason, 1993). Strictly speaking, although the residual errors are not normally distributed, the parameter estimates from ML or ReML are still consistent and asymptotically unbiased. However, the asymptotic standard errors are incorrect and the corresponding statistical inferences are not trustworthy. Furthermore, these problematic consequences are not completely vanquished even if a larger sample size is provided (Goldstein, 1995; Maas & Hox, 2004).

Previous meta-analyses have shown that normality assumption is not always satisfied in MB studies (e.g., Parker, 2006; Shadish, 2014; Shadish & Sullivan, 2011; Smith, 2012, Solomon, 2014). For example, Parker (2006) investigated the normality of single-case data where 166 published data sets were analyzed with the Shapiro-Wilk test (Shapiro & Wilk, 1965). Results indicated that a full 51% (N = 85) of these 166 data sets failed to meet the normality assumption (Parker, 2006). In addition, a meta-analysis of school-based single-case studies reported that the skewness of analyzed study observations ranged from .46 to 2.89 and the corresponding kurtosis ranged from .49 to 1.57 (Solomon, 2014).

Non-normal data may occur in MB studies due to various reasons. Non-normality can be observed if the scales of measurement are not continuous variables including counts, proportions or percentages. In MB studies, scales of measurement vary from study to study depending on the researchers' interest. For example, if a researcher is interested in the play behavior of toddlers with disabilities, then a count of play actions could be the target outcome variable (DiCarlo &

Reid, 2004). If a researcher is interested in the academic and social participation of students with disabilities during multiple sessions or trials, then the proportion of students' initiations of social interactions with the teacher or other students could be the target outcome variable (Hunt, Soto, Maier & Doering, 2003). Although the measurement scales are continuous, the distribution may be still non-normal if the participant's measurement occasions included ceiling/floor effects or outliers (Hox et al., 2010; Langford & Lewis, 1998). In MB studies, ceiling/floor effects or outliers may occur due to either a momentary or temporary event effect. For example, a momentary event effect occurs if a participant of an intervention study experiences a personal problem at home at a certain time point and the observed outcome at the next time point is influenced by the event effect, either positively or negatively (Ferron et al., 2014).

If non-normality in MB studies were observed, one could use generalized linear (mixed) models assuming either Poisson or binomial distributions as underlying population distributions to fit the non-normal scale observations (Shadish, 2014). However, this approach includes some limitations. For example, fitting more complex models can increase the complexity of the estimation which can create problems with small sample sizes (Shadish, Kyse, & Rindskopf, 2013). In addition, in practical situations, the researcher may not know the correct underlying population distribution with any confidence (Shadish, Zuur, & Sullivan, 2014). Lastly, if MB studies were meta-analyzed, it is more challenging to put different scales of effect size measures on the same metric across studies.

Alternatively, one could use multilevel models assuming robustness of the models to the violation the normality assumption. It may not be an ideal approach if severe violation of normality led to an inaccurate interpretation of the results. However, up to date, the information regarding the robustness of multilevel models is limited and no guidelines or recommendations

with respect to skewness and kurtosis of MB data are provided. Thus, it is important to investigate the degree to which violation of normality assumption in MB studies can be handled with multilevel models. In addition, given that non-normal MB data is a potential threat and could possibly lead to inaccurate statistical inferences, it is important to investigate an alternative modeling methods, which may be robust to non-normal MB data. From a practical perspective, comparing traditional and alternative modeling methods can provide better insight and practical solutions for applied researchers about how to deal with a violation of the normality assumption.

Bayesian Estimation and Inference

One alternative approach for handling non-normal data in multilevel modeling is Bayesian estimation (Gelman, Carlin, Stern & Rubin, 2014). The Bayesian approach conceptually and methodologically differs from likelihood-based estimation methods. In Bayesian modeling, researchers can specify prior knowledge about the model parameter and this prior knowledge is often expressed as a probability distribution, also known as *prior distribution*. For example, in multilevel modeling, prior distribution can be specified for each fixed effect coefficient and variance component in the model before the parameter estimation. In general, the prior distribution is determined based on a researcher's belief or prior knowledge and it could significantly affect the precision and inference of the parameter estimate. This significantly-influencing prior distribution is generally called an *informative prior distribution.* If a researcher does not have any prior knowledge of the model parameter, then a "non-informative" or flat distribution should be specified as the prior distribution. This is also called a *non-informative prior distribution.* Either an informative or non-informative prior distribution is combined with the likelihood probability of the model to create the *posterior distribution*, which is a distribution

of the product of prior and likelihood probabilities for the model parameter. Unlike classical statistics, the Bayesian perspective considers a model parameter as a random variable rather than a fixed parameter so that parameter possesses its own probability density function. The posterior distribution represents the probability density function for the model parameter and it can be comparable to the concept of a sampling distribution in classical statistics. The model parameter estimate can, then, be obtained from simply taking either the expected value of the posterior distribution (i.e., Expected a Posteriori [EAP] estimate) or mode of the posterior distribution (i.e., Maximum a Posteriori [MAP] estimate).

Bayesian estimation has several advantages over likelihood-based estimation. First, the Bayesian method does not require a large sample size to obtain an accurate parameter estimate. Bayesian estimation often works well when the sample size is relatively small because it takes advantage of the prior distribution on the parameter of interest (Gelman et al., 2014). For example, if a researcher has a basic idea such as the possible minimum or maximum values for the parameter based on previous studies, then putting the prior distribution that contains the certain minimum and maximum values would prevent an estimation of the extreme values beyond these boundaries from occurring. In multilevel modeling, when the variance components are estimated, it is generally known that variances cannot be below zero, thus, the appropriate prior distribution would be a positive-valued distribution such as an inverse gamma distribution or an inverse chi-square distribution (Gelman, 2006). Previous simulation studies also have shown that using Bayesian estimation with an appropriate prior distribution produced reasonable fixed effect and variance component estimates with relatively smaller sample sizes than ML (e.g., Browne & Draper, 2000, 2006; Browne et al., 2002).

Another advantage of Bayesian estimation is the relative ease of the computational procedure. In principle, the likelihood-based estimation methods require the analytic form of the first- and second-order partial derivatives with respect to parameters of the model to obtain the parameter estimates and their standard errors. For example, in multilevel modeling, the standard error estimate is obtained from an inverse of Hessian matrix, which requires the second derivative of the likelihood function (i.e., Fisher Information). The derivative forms often get extremely complicated when the fitting model includes a large number of parameters. In addition, likelihood-based methods use an iterative procedure, which may cause a convergence problem when complex models are fitted. On the contrary, Bayesian estimation does not require a complex analytic form of a derivative function to obtain the parameter and standard error estimates. Rather, the Bayesian approach uses a sampling method, also known as Markov chain Monte Carlo (MCMC). Thus, the posterior distribution of parameter can be obtained using MCMC sampling from a product of prior distribution and likelihood function. This MCMC sampling allows researchers to take a sample of any size from the posterior distribution of the parameters of the model. The MCMC sampling method makes it easy to estimate any function of parameters even though a large number of parameters are involved in the model (only the sampling time gets longer). Note that there are a number of MCMC sampling methods proposed in the statistical literature. Gibbs sampling and Metropolis-Hastings (MH) sampling methods are the most commonly used algorithms in the applied literature (e.g., Spiegelhalter, Thomas, Best, Gilks, & Lunn, 2003). Generally speaking, Gibbs sampling draws samples from joint probabilities of prior and likelihood distributions to create posterior distribution of a parameter with the assumption that other parameters are unknown. This drawing procedure continues sequentially for all the parameters until the number of iterations reaches the maximum. MH

sampling, however, has the ability to draw samples from an arbitrary functional form of a distribution with an unknown scale so that it can compare the probability of the drawn sample with the joint probabilities of prior and likelihood distributions to decide whether the sample is accepted or not. Theoretically, both sampling algorithms asymptotically create the posterior density distribution of a parameter when the sampling algorithm shows convergence (Gelman et al., 2014).

Bayesian inference of the parameter estimates is also an advantage of using Bayesian estimation. In classical statistics, statistical inferences such as the *p*-value or confidence interval are often computed from the theoretical sampling distribution. The sampling distribution for a parameter, however, is based on the repetition of samples of a fixed quantity and the probability interpretation for the fixed quantity is not exactly the same as the probability interpretation for a parameter, which is assumed to have its own probability distribution (Rindskopf, 2014; Swaminathan, Rogers, & Hornor, 2014). The probability statement for the parameter makes more sense when the parameter is considered as a random variable rather than a fixed parameter. In Bayesian statistics, statistical inference is made in accordance with the posterior distribution, which is considered as the probability of the parameter (Gelman et al., 2014). Note that, again, the debate between classical and Bayesian perspectives is not a primary focus of this study so a detailed argument is not be further discussed. Instead, the literature review is focus more on current literature for Bayesian estimation and its implementation to single-case data multilevel modeling.

Lastly, a number of statistical software programs have been developed for Bayesian estimation and are available for applied researchers, such as WinBUGS, SAS PROC MCMC, JAGS, MLwiN, Mplus and R package MCMCglmm. These software programs are flexible

enough to fit a wide variety of statistical models including single- or multi-level models, linear or non-linear models (i.e., generalized linear models), and homogeneous or heterogeneous variances models. Several methodological studies also have demonstrated how to conduct Bayesian analysis with more complex statistical models, providing the program codes, output results from the analysis, and their interpretation. For example, Rindskopf (2014) provided illustrations of Bayesian data analysis for single-case data using linear and non-linear models. He demonstrated the empirical data analysis using the software program WinBUGS and provided the relevant codes, output of the results, and a detailed interpretation of the results. In addition, Swaminathan et al. (2014) illustrated Bayesian analysis for an effect size measure for single-case data. They also used the software program WinBUGS and provided the code with detailed explanations.

Bayesian Modeling and MB Studies

Over the last decade, Bayesian method has been exclusively integrated with multilevel modeling (e.g., Brown & Draper, 2000, 2006; Browne et al., 2002; Baldwin & Fellingham, 2013) and applied in the context of single-case research (e.g., Baek, 2015; Rindskopf, 2014, de Vries, & Morey, 2013; Shadish, et al., 2013; Swaminathan et al., 2014). For example, Rindskopf (2014) argued that Bayesian multilevel modeling has a number of advantages over likelihood-based estimations in analysis of data from studies such as MB studies because: a) it is more suitable for analyzing studies with small sample sizes, b) it is more interpretable than the results from classical statistics, and c) when a parameter is estimated, it takes into account the uncertainty about all other parameters so that larger standard errors accurately reflect the totality of the uncertainty about the model parameters. Previous simulation studies have also shown its

effectiveness in multilevel modeling. Studies found that fixed effects were well-estimated with small sample sizes and inferences were reasonably accurate (e.g., Baldwin & Fellingham, 2013, Baek, Petit-Bois, & Ferron, 2014). For example, Baldwin and Fellingham (2013) found that the treatment effect estimate was unbiased when the level-2 sample size was as few as 8, and the corresponding coverage rate was equally accurate as ReML using Kenward-Roger adjustment. However, similar to ReML, the variance component estimates were still substantially biased and inferences were not accurately estimated (Baldwin & Fellingham, 2013). Furthermore, previous studies compared estimation accuracies of traditional likelihood-based estimations (ML and ReML) and Bayesian estimations incorporating different types of the prior distributions for the within-series multilevel modeling of MB studies. They found that both likelihood-based and Bayesian estimations recovered the treatment effect estimates without biases and confidence interval coverages were close to the nominal level (e.g., Moeyaert et al., 2016). However, note that previous Bayesian modeling of MB studies were mainly focused on the within-series model and as yet, Bayesian estimation efficacy for the between-series model has not been investigated.

Bayesian estimation also has been implemented to accommodate many possible complications in multilevel modeling with MB studies, previously. For example, Baek (2015) investigated multilevel modeling with the heterogeneous variances across participants, and the convergence issue occurred when the complex model was fitted with ReML. The convergence issue was then resolved when Bayesian estimation was implemented by specifying the prior distribution to each variance component of the model (Baek, 2015). Moreover, Gelman et al., (2014) and Rindskopf (2014) suggested that if the dependent variables are not normally distributed due to outliers or different underlying distributions, then specifying a thicker-tailed prior distribution can accommodate the complex situation. However, up to date, no research has

examined the efficacy of Bayesian modeling for non-normal MB data. Also, studies have not been conducted that compare of ReML and Bayesian estimation for both within- and between-series models under the violation of normality assumptions. To fill this gap, the current study is aimed to investigate the efficacy of alternative Bayesian modeling for non-normal data in MB studies. A detailed description of Bayesian modeling such as the prior specification for the model, initial values, number of iterations, convergence criteria and related statistical software programs are further discussed in the next chapter.

Summary

A single-case study is a type of experimental study used to investigate the effect of an intervention or treatment for case-specific observations. The most popular design in single-case studies is the multiple-baseline (MB) design (Shadish & Sullivan, 2008). MB design studies are particularly different from other group experimental studies since a relatively small number of participants is involved in a study. Also, an MB study has stronger internal and external validity than a basic AB design in single-case studies because it allows researchers to examine the treatment effect comparing not only within-participants' observations from baseline phase to treatment phase but also between-participants' observations at certain time points. Several statistical methods were proposed to analyze the treatment effect in MB studies and two statistical models were recently proposed for analyzing MB studies: within-series and between-series models. Previous research has shown that the within-series model outperforms the between-series model when assumptions are satisfied. When assumptions are violated due to event effects or model misspecification, the between-series model produced less biased estimates of the treatment effect than the within-series model (Ferron et al., 2014). However, the

robustness of the within- and between-series models for non-normality still remained questionable and little research has investigated this issue. Given that meta-analyses reported that data from single-case studies, including MB studies, tend to be non-normal due to the scales of measurement, ceiling/floor effects, or outliers, it is worthwhile to compare various modeling approaches under violation of the normality assumption. Therefore, the purpose of this study is to investigate the robustness of various models for MB studies when the normality assumption is violated. This study includes Bayesian estimation and inference as an alternative approach because in theory it has several advantages over maximum likelihood estimation with respect to non-normality.

CHAPTER THREE: METHODS

The methods section describes simulation design, data generation, fitting models, and dependent variables of the simulation study.

Simulation Design

A Monte Carlo study was conducted to empirically address the issues of violation of the normality assumption in MB studies. The simulation design included three design factors (number of measurement occasions, number of participants and population treatment effects) and non-normality factors (skewness and kurtosis). The three design factors were varied with a) the number of measurement occasions having values of 10, 20, and 40, b) the number of participants having values of 4, and 8, and c) population treatment effect value having values of 0 and 1. Two non-normality factors were varied with a) skewness of the level-1 residuals having values of 0, 1, 2, and 3 and b) kurtosis of the level-1 residuals having values of -1, 0, 1, 2, and 4. Crossing all the simulation design factors resulted in a total of 3 (the number of measurement occasions) x 2 (the number of participants) x 2 (population treatment effects) x 4 (the level of skewness) x 5 (the level of kurtosis) = 240 simulation conditions. For each condition, 3000 data sets were generated. The number of replications was chosen based on the previous simulation studies with single-case studies (e.g., Ferron et al., 2009, 2010). The replicated data were, then, analyzed with 4 different approaches (Models 1 - 4 as described above). The parameters of the within- and between-series models were estimated using ReML estimation with Kenward-Roger adjustment

(Models 1 and 2) and Bayesian estimation (Models 3 and 4). Data generation procedure was conducted using SAS/IML program (SAS Institute, 2014). ReML with Kenward-Roger estimation was accomplished using the SAS MIXED Procedure and Bayesian estimation was accomplished using the SAS MCMC Procedure.

Design Factors

In the current study, three levels of measurement occasions (10, 20 and 40) and two levels of participants (4 and 8) were considered. The number of measurement occasions was chosen based on meta-analyses results for single-case research studies. A meta-analysis of 85 single-case studies found that 25 studies had fewer than 11 measurement occasions, 37 studies had between 11 and 29 measurement occasions and 23 studies had more than 29 measurement occasions (Swanson & Saches-Lee, 2000). Another meta-analysis of single-case studies also found that a median number of measurements were 20. This meta-analysis further identified that 90.6% of the participants had 49 or fewer measurement occasions (Shadish and Sullivan, 2011). In accordance with these meta-analyses results, previous simulation studies with single-case research have included that the number of measurement occasions within a participant varied from 10 to 40 (e.g., Moeyaert et al., 2013a, 2013b, 2014).

Similarly, the numbers of participants were chosen based on previous single-case studies (e.g., Kazdin & Kopel, 1975). Traditionally, single-case studies include small numbers of participants and the typical number of participants has been four (Ferron et al., 2010; Kazdin, 2011). In addition, Shadish and Sullivan (2011) meta-analyzed 809 single-case studies and found that the number of participants per study ranged from 1 to 13. Furthermore, Farmer, Owens,

Ferron and Allsopp (2011) found that in 93% of the surveyed multiple-baseline studies the number of participants was 7 or less.

In order to investigate the statistical power and Type I error of the treatment effect estimate across fitting models, two different population treatment effects were included (0 and 1). Note that when there was no true treatment effect in the data generation step, Type I error were computed, whereas when there was a true treatment effect, statistical power were computed across simulation conditions.

Non-normality Factor

Non-normality of level-1 error variance was created by manipulating skewness and kurtosis of the population distribution. To investigate the impact of non-normality, skewness and kurtosis were allowed to vary by equally spacing values from commonly observed ranges in single-case studies. Skewness of the level-1 errors was varied from 0 to 3 (i.e., 0, 1, 2, and 3) and kurtosis was varied from -1 to 4 (i.e., -1, 0, 1, 2, and 4). The direction of the skewness was set as positive across simulation conditions because one of the major focuses in the study is the violation of symmetric assumption for the multilevel modeling rather than the direction of the skewness.

The ranges of the skewness and kurtosis were chosen based on the previous studies (e.g., Owens & Farmer, 2013; Solomon, 2014). A simulation study conducted by Petit-Bois et al. (2013) investigated the non-normality of level-2 and level-3 error structure in meta-analytic multilevel modeling for MB studies and they included skewness from 0 to 1.75 and kurtosis from 0 to 3.75. Owens and Farmer (2013) included non-normality of level-1 and level-2 error structure varying the skewness from 0 to 1 and kurtosis from -1 to 3.75. A meta-analysis also supports that

skewness of analyzed study observations ranged from .46 to 2.89 and the kurtosis ranged from .49 to 1.57 (Solomon, 2014). In addition, the author further conducted a preliminary survey investigating the ranges of skewness and kurtosis of the level-1 residuals from MB studies published in *Journal of Applied Behavior Analysis (JABA)* from 2014 to 2016. A total of 20 datasets excluding binary dependent variables were collected and fitted with the two-level within-series model. Because the variance component estimates in the between-series model contain both level-1 and level-2 error variances, the within-series model is more appropriate to obtain level-1 residual distributions. Skewness and kurtosis of the level-1 residuals of the model were computed individually and they ranged from -0.71 to 1.91 for skewness and -1.07 to 3.01 for kurtosis, respectively.

Based on these previous investigations, the current study examined the skewness and kurtosis of level-1 errors across the range of values previously investigated and observed, but varied the values more systematically and in smaller increments. This approach could provide a detailed guideline for applied researchers and practitioners about how to deal with degree to which skewness and kurtosis of MB studies.

Data Generation

Data Generation Models

The data generation took place in two stages. First normally distributed data were generated and then these data were transformed to induce desired levels of skewness and kurtosis. The initial data generating model is described as follows.

Level-1 Equation:

$$Y_{ij} = \beta_{0j} + \beta_{1j}Phase_{ij} + \beta_{2j}Time_{ij} + \beta_{3j}(Time_{ij} * Phase_{ij}) + e_{ij} \qquad (13)$$

Level-2 Equation:

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (14.1)$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (14.2)$$

$$\beta_{2j} = \gamma_{20} + u_{2j} \quad (14.3)$$

$$\beta_{3j} = \gamma_{30} + u_{3j} \quad (14.4)$$

Level-1 errors were initially generated to be independent, normal, and homogeneous across phases and participants with a mean of 0 and a variance of 1. The variance-covariance matrix for level-2 errors (u_{0j} u_{1j}, u_{2j}, and u_{4j}), was an uncorrelated diagonal matrix. That is, $\boldsymbol{\Sigma_u}$ = *diag*(τ_{00}, τ_{11}, τ_{22}, τ_{33}). The uncorrelated diagonal matrix was assumed because more complex level-2 error structure could yield more biased level-2 error variance estimates given that level-2 error variance estimates are generally biased in single-case data (Moeyaert et al., 2013a, 2013b, 2014). Also a previous study showed that misspecification of level-2 error structure has a minimal impact on the treatment effect estimate (Moeyaert et al., 2016). For the condition where power of the treatment effect was investigated, the population parameter values for fixed effect coefficients, γ_{00}, γ_{10}, γ_{20}, and γ_{30} were set as 0, 1, 0, and 0, respectively. Similarly, for the condition where Type I error of the treatment was examined, the data generation parameter values for the all fixed effect were set as 0. These parametrization of the model implies that the time trend effect is not included in the simulated data. These parameters were chosen because the primary focus of the current study is of the treatment effect estimation accuracy and inference. Population parameter values for level-2 variance components, τ_{00}, τ_{11}, τ_{22}, and τ_{33} were set as .50, .50 .00, and .00, respectively. This parametrization indicates that there are variations across participants in baseline observations and treatment effect sizes but not in time trend effects

in baseline and treatment phases. The population parameter values for level-2 variance components were chosen based on previous simulation studies with single-case studies (e.g., Ferron et al., 2009, 2010; Moeyaert et al., 2013a, 2013b, 2014, 2016).

Data Generation Steps

The simulated data were generated using SAS/IML program. For each replication, the values for level-1 and -2 errors were first randomly generated from standard normal distribution, individually, using RANNOR function implemented in SAS/IML. Note that number of randomly generated values for level-1 error variances was *I* x *J*, where *I* is the number of measurement occasions, and *J* is the number of participants. Also, the number of randomly generated values for level-2 error variances was 4**J* and each distinct value was replicated by *I* times. 4**J* number of random values were created because the number of fixed effect parameters in the data generation model was four, and *J* number values were needed to create variations across participants. This strategy created each parameter's level-2 error variation across participants (for each parameter, τ_{00}, τ_{11}, τ_{22}, and τ_{33}, respectively) and within-participants variation is only affected by level-1 error variance values.

Furthermore, the *time* variable was created as a sequence of integers corresponding to the session number, the *phase* variable as the baseline or treatment phase indicator (coded 0 = baseline, 1 = treatment) and the *phase*time* variable as the treatment observation time interaction. The variable *phase*time* was created by simply multiplying two variables *phase* and *time.* Note that the *phase* variable was created to mimic MB studies such that each participant had the different intervention time points. For example, for the condition where the number of participants was four and the number of measurement occasions was 10, the

first, second, third and fourth participant's *phase* variable became 1 when their observations were at 4, 5, 6, and 7th time point, respectively so that all phases had at least 3 observations. Similarly, for the condition where the number of participants was eight and the number of measurement occasions was 10, pairs of participants' *phase* variables were created. That is, first and second participants had the same intervention time point, third and fourth participants had the same intervention time point and so on. When the numbers of measurement occasions were 20 and 40, the intervention phases started at 5, 8, 11, and 14th observations and 10, 16, 22, and 28th observations for the first, second, third and fourth participants, respectively. It is also important to note that variables *time* and *phase*time* were created by the group-centering approach, meaning each participant's *time* was 0 when treatment phase started. This group-centering approach allows creating data sets where the treatment effect is an immediate shift in level at the initial treatment observation for each participant.

After error variance values and predictor variables were generated, two different data generation routes were taken depending on the simulation conditions to generate a dependent variable Y_{ij}. For the condition where normality assumption was not violated, all error variance values were combined with predictors parameters (γ_{00}, γ_{11}, γ_{22}, and γ_{33}) and their variables were used to create the dependent variable Y_{ij} Equation 13 as described above. For the condition where normality assumption was violated, a series of level-1 error values for each participant was manipulated using Fleishman's power transformation method (Fleishman, 1978). Using Fleishman's method, certain degrees of skewness and kurtosis can be achieved. A detailed explanation about the power transformation method is discussed in later section. Once the level-1 error values are manipulated as desired, the

simulated dependent variable Y_{ij}, were again generated using Equations 13 and 14.1 - 14.4. Saving the id variable for both level-1 and level-2 units, total six variables were created for each simulated data set; Y_{ij}, *idlevel1*, *idlevel2, phase, time,* and *phase*time.*

For the between-series model parameterization, additional four dummy variables (P_{1ij}, P_{2ij}, P_{3ij}, and P_{ij}) were created in the simulated data sets. P_{1ij}, P_{2ij}, P_{3ij}, and P_{ij} were indicators where the values became 1 if observations were used for the between-participant comparison and 0 otherwise. More specifically, P_{1ij} = 1 if i^{th} observation for j^{th} participant was used for the first time point between-participant comparison. Similarly, P_{2ij} = 1 if i^{th} observation for j^{th} participant was used for the second time point between-participant comparison. And, P_{3ij} = 1 if i^{th} observation for j^{th} participant was used for the third between-participant comparison. Finally, P_{ij} = 1 if P_{1ij} = 1 or P_{2ij} = 1 or P_{3ij} = 1, 0 otherwise.

Fleishman's Power Transformation Method

Fleishman (1978) proposed a power method to generate non-normal distribution data. The proposed method allows manipulating skewness and kurtosis of the standard normal distribution using the powers of polynomial equations. The polynomial equation is given as follows.

$$Y = a + bX + cX^2 + dX^3 \quad (15)$$

where Y is transformed non-normal variable with specified population skewness and kurtosis, X is standard normal variable and *a, b, c* and *d* are constants needed for transforming the standard normal distribution to non-normal distribution with specified degrees of skewness and kurtosis. Note that $a = -c$. In Fleishman's article (1978), the coefficient values (*a, b, c* and *d*) are provided to generate the non-normal distribution with

specified population skewness and kurtosis. For example, if one desires to generate a distribution with population skewness = 2 and kurtosis = 3, then the constants values are as follows.

$$b = 0.92966052480111$$

$$c = 0.39949667453766$$

$$d = -0.03646699281275$$

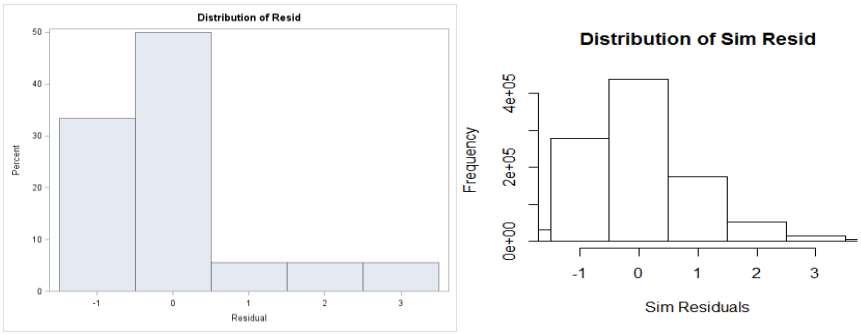
Also, note that standard normal distribution can be obtained when $b = 1$, $c = 0$, $d = 0$.

Fleishman's transformation coefficients for the current study conditions were computed using SAS/IML program. The computed coefficients are shown in Table 2. Note that once the transformation was made with the Fleishman's method, the values were standardized to set mean of zero and standard deviation of one.

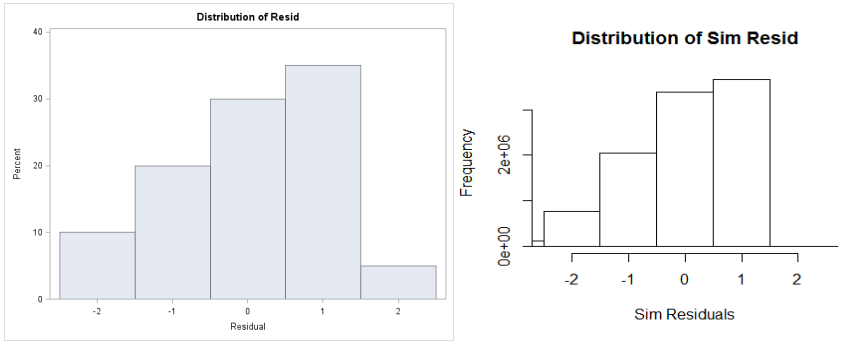
Table 2.
Fleishman Transformation Coefficients for Various Degrees of Skewness and Kurtosis

Skewness	Kurtosis	*a*	*b*	*c*	*d*
0	-1	0	1.22101	0	-0.0802
0	0	0	1	0	0
0	1	0	0.90298	0	0.03136
0	2	0	0.83566	0	0.05206
0	4	0	0.73738	0	0.08093
1	-1	0.38757	-3.0507	-0.3876	2.56235
1	0	-0.2394	1.08828	0.23938	-0.0422
1	1	-0.191	1.01749	0.191	-0.0186
1	2	-0.1472	0.90476	0.14721	0.02386
1	4	-0.117	0.77659	0.11698	0.0655
2	-1	-0.2341	-49.851	0.23413	8.36508
2	0	-14.858	-26.684	14.858	4.18199
2	1	4.00552	-4.1548	-4.0055	-0.8493
2	2	-1.5751	0.81684	1.57508	-0.1344
2	4	-0.3389	0.93083	0.33887	-0.0084
3	-1	-0.1207	1.17205	0.12073	0.29789
3	0	634.454	250.927	-634.45	-13.804
3	1	-0.2651	0.6859	0.26508	0.08867
3	2	1.22953	4.79678	-1.2295	-0.7268
3	4	-0.7709	-4.0862	0.77087	0.49931

(a) Residual Skewness = 1.91 & Kurtosis = 3.01 (Edwards et al., 2015)



(b) Residual Skewness = -.60 & Kurtosis = -.27 (Himle & Wright, 2014)



(c) Residual Skewness = .55 & Kurtosis = .69 (Washington, Banna, & Gibson, 2014)

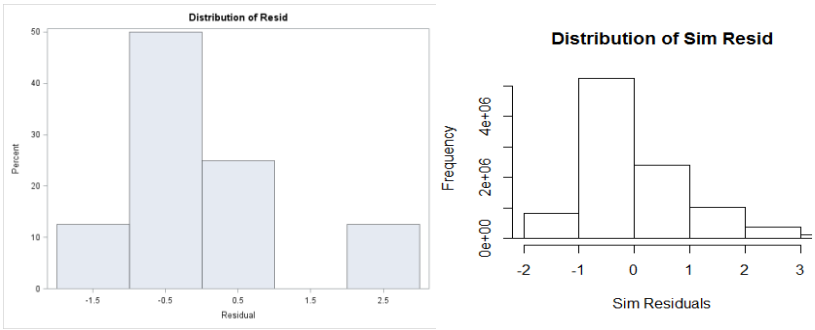


Figure 5. Residual distributions from multiple studies (left) and simulated data (right).

Also, for the illustration purpose, Figure 5 shows the histograms of residuals from the preliminary survey data and simulated data to show replicability of simulated data to real data residual distributions. The preliminary survey data were fitted with two-level within-series model

and residuals were computed. Also, the computed residuals were standardized to make the same scale as the simulated data. For the simulated data, 1 million random samples were drawn from standard normal distribution and then transformed using Fleishman's equation to mimic the skewness and kurtosis of the survey residual distribution.

Fitting Models

Within-Series Model

The equation for the within-series model is described as follows.

$$Y_{ij} = \gamma_{00} + \gamma_{10}Phase_{ij} + \gamma_{20}Time_{ij} + \gamma_{30}\left(Time_{ij} * Phase_{ij}\right)$$
$$+u_{0j} + u_{1j}Phase_{ij} + u_{2j}Time_{ij} + u_{1j}\left(Time_{ij} * Phase_{ij}\right) + e_{ij}, \quad e_{ij} \sim N(0, \sigma^2) \quad (16)$$

Note that the fitted within-series model is equivalent to the data generation model. The level-1 error variance was assumed to be independent and homogeneous across phases and participants. Level-2 error structure was assumed to be an uncorrelated diagonal matrix.

$$\Sigma_u = diag(\tau_{00}, \tau_{11}, \tau_{22}, \tau_{33})$$

Between-Series Model

The equation for the between-series model is described as follows.

$$Y_{ij} = \beta_1 P_{1ij} + \beta_2 P_{2ij} + \beta_3 P_{3ij} + \beta_4 P_{ij} * Phase_{ij} + e_{ij}, \quad e_{ij} \sim N(0, \sigma^2) \quad (17)$$

The between-series model contains four parameters using three comparison time-point observations across participants and one pooled treatment effect parameter. The between-participant comparison time points were fixed at the third observations after the treatment phase started. As discussed earlier in chapter two, separate time points baseline observation estimates allows the model to have time trend effect for baseline phases.

Residual variance of the model, σ^2, was assumed to be independent and homogeneous across phases and participants. Note that the between-series model residual σ^2 includes both within- and between-participant variations.

Parameter Estimation

Parameters of within- and between-series models were estimated using ReML and Bayesian estimation methods. When ReML was used to estimate parameters of the model, the Kenward-Roger approach was used to compute adjusted standard error and degrees of freedom for small sample size. SAS PROC MIXED was used to estimate the parameters of the models using ReML with Kenward-Roger adjustment.

Prior Distribution

Prior specification for parameters in the model is an important step in Bayesian estimation. Based on previous Bayesian multilevel modeling studies, the following prior distributions were specified for the within- and between-series models. For the within-series model, prior distributions for fixed effect parameters γ_{00}, γ_{11}, γ_{22}, and γ_{33} were assumed to be the normal distribution with zero mean and 1×10^{10} variance. In addition, prior distributions for level-1 and level-2 error structure were assumed to be the inverse-Wishart distribution with degrees of freedom parameter ν = 0 and expected parameter *V* was a 4 x 4 identity matrix. Note that both normal and inverse-Wishart distributions are considered as conjugate priors for the fixed effects and random effects, respectively, meaning the posterior distribution can be theoretically derived from the prior distribution and likelihood function (Gelman et al, 2014). These prior specifications were also chosen based on the previous Bayesian multilevel modeling

studies. Previously, dispersed distributions have been heavily used as prior distributions for both fixed effect and variance components (e.g., Baldwin & Fellingham, 2013; Browne & Draper, 2006; Gelman, 2006; Moeyaert et al., 2016; Rindskopf, 2014).

However, it is important to note that although Bayesian multilevel modeling introduces the prior distribution, which theoretically reasonable for the variance component parameter, it was found that the variance components estimates are still substantially biased if the sample size is relatively small (Baldwin and Fellingham, 2013; Moeyaert et al., 2016). Limited research, yet, have found more effective and precise prior distributions for the variance components for small sample size conditions in Bayesian multilevel modeling.

For the between-series model, similar prior distributions as the within-series model for the parameters were specified. Prior distributions for the parameters, β_1, β_2, β_3 and β_4 were assumed to be the normal distribution with zero mean and 1×10^{10} variance, respectively. Also, the prior distribution for the variance component σ^2 was assumed to be the inverse-Wishart distribution with the same parameterization as the within-series model prior specification. Note that because the assumptions of independence and homogeneity across phases and participants hold for the level-1 error structure in the between-series model as well, the inverse-Wishart distribution became univariate inverse-gamma distribution.

Convergence Criteria

For the current simulation study, the convergence rates for both ReML and Bayesian estimations were recorded and summarized. The convergence rates were computed as proportions of replications in which estimations were reaches the convergence.

Unlike ReML, Bayesian estimation has several criteria for determining convergence of the sampling procedure. Once MCMC samples reach a certain degree of stability across iterations, posterior distribution samples are considered as converged. MCMC convergence can be checked with a) Kernel density plots of samples, b) history or trace plots of the mixing procedure, c) autocorrelation between adjacent posterior samples and d) statistical diagnostics. Various statistical diagnostics are developed in statistical literatures including $\hat{R}$ (Gelman & Rubin, 1992), Geweke test (Geweke, 1992), Heidelberger-Welch stationary and half-width tests (Heidelberger & Welch, 1983) and Raftery-Lewis test (Raftery & Lewis, 1992). In current study, Geweke test was used to evaluate the convergence rates. Note that Geweke test evaluates the convergence of the Markov chain samples by comparing means from the early and latter part of the Markov chain. Significant differences between two parts of the Markov chain samples indicate that MCMC procedure does not reach to the convergence.

A preliminary simulation study was conducted to examine the sufficient number of iterations for the convergence. The study indicated that the Markov chain samples were stable after a burn-in period between 5,000 and 10,000 iterations. In order to achieve the convergence in the final simulation across conditions, 100,000 iterations with the 10,000 burn-in period were conducted and only every 25th draws after burn-in period were kept (thinning =25). This strategy was chosen based on the previous Bayesian estimation studies (e.g., Moeyaert et al., 2016). For the current study, the convergence status for each replication was recorded for both ReML and Bayesian estimations. The convergence rates for both estimation methods are reported across simulation conditions.

Estimation

Once the MCMC procedure reaches the maximum number of iterations, the parameters of the models were, then, estimated by taking a mean of the posterior samples. One could question whether taking a median of the posterior samples is more reasonable than a mean for the variance components because the posterior density of the variance component tends to be a right-skewed distribution. However, across the simulation conditions, minimal differences between mean and median were found. Therefore, to keep the consistency of the results, only mean statistics of the posterior samples were reported for the variance components.

Dependent Variables

Simulation results are analyzed with various criteria. For the accuracy of parameter estimation of the within- and between-series models, the following statistics were used to analyze the accuracy of the estimation methods:

$$Bias = \frac{\sum_{i=1}^{R} \widehat{\beta_i} - \beta}{R} \tag{18}$$

$$Relative\ Bias = \frac{\sum_{i=1}^{R} \frac{\widehat{\beta_i} - \beta}{\beta}}{R} \tag{19}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{R} \left(\widehat{\beta_i} - \beta\right)^2}{R}} \tag{20}$$

R denotes the total number of replications, β denotes population (generating) parameters and $\widehat{\beta_i}$ represents estimated parameters for i^{th} replication for the multilevel modeling. Note that bias and relative bias were computed for each replication then averaged across replications. RMSE was computed by taking the sum of square differences between population and estimated parameters for replications then averaged across replication (mean square error). The final RMSE was

obtained by taking the square root of the mean square error. Previous research noted that relative bias less than .05 can be considered as an acceptable bias for the fixed effect estimates and .10 for the variance component estimates (Hoogland & Boomsma, 1998).

As a measure of statistical inference, interval estimate coverage rates, interval estimate width, and statistical power/Type I error of the treatment effect parameter were computed. When parameters were estimated with ReML, confidence interval (CI) coverage rates, CI widths, and power/Type I error were computed as traditional fashion. That is, SEs and degrees of freedom were computed with Kenward-Roger method and CIs were obtained using those quantities. CI coverage rates were, then, obtained from taking proportions of replications in which population parameters were inside of the computed CI. CI widths were also computed taking the difference from upper bound to lower bound of CIs per replications then averaged across replications. Statistical power was also computed for treatment effect where the *p*-value was less than nominal level of significance. Empirical statistical power was, then, obtained from the proportion of replications in which *p*-value was less than the nominal level. The nominal level of significance was set as .05. Type I error rates were also computed similarly as statistical power when true treatment effect was zero.

When Bayesian estimation was used, highest posterior density (HPD) was used to compute interval estimates for the treatment effect. Note that HPD is essentially different than credible interval in Bayesian inference. The *credible interval* is generally computed taking 2.5% and 97.5% quartiles of the overall posterior distribution as lower and upper bounds, respectively. However, the HPD interval is computed taking the 2.5% quartile and 97.5% quartile of the highest posterior sample density for the lower bound and higher bound HPD intervals, respectively. These credible and HPD intervals do not always produce the same intervals. If the

posterior distribution is not symmetrical or bimodal shape, then credible and HPD intervals are substantially different. In the current study, the HPD coverage rate and its interval width were computed the same fashion as ReML computes CI coverage rates and CI widths. Statistical power and Type I error rates for Bayesian estimation were computed as proportions of replications in which the HPD intervals contained 0. This computation is comparable to the ReML approach.

Because 3000 datasets for each condition were simulated, the coverage proportions for both ReML and Bayesian interval estimates should be estimated relatively accurately. To evaluate the coverage proportions, an acceptable range of the coverage estimates was computed using the standard error of the coverage probability, p (Burton, Altman, Royston, & Holder, 2006). The standard error equation is

$$SE(p) = \sqrt{[p(1-p)]/B} \tag{21}$$

where p represents the nominal coverage probability of .95, and B is 3000 (the number of replications per condition), resulting in a range from .942 to .958 for acceptable coverage estimates. The estimation accuracy for the parameters other than treatment effect such as coefficients for intercept, time, and interaction and level-2 and level-1 error estimates in the within-series model and level-1 error estimate in the between-series model were analyzed by computing only bias and RMSE across replications.

Analysis of Dependent Variables

Results tables of the dependent variables across simulation conditions are presented in the results section. To analyze the variation in outcomes as a function of the simulation design and non-normality factors, multi-way univariate ANOVAs were conducted on biases, relative biases,

RMSEs, coverage rates, coverage widths, and statistical power/Type I error. The ANOVA analyses are appropriate to determine the simulation condition effects and their effect size measures. Eta-square (η^2) statistics were additionally computed based on the ANOVA results to investigate the effect size of the design and non-normality factors. Eta-square is often used to compute the amount of explained variance for each factor, along with significant tests. Eta-square was computed using the proportion of variability of each dependent variable that is associated with each of the effects in the simulation conditions. The ratio of the effect variance (SS_{effect}) to the total variance (SS_{total}) yields the eta-square statistics.

$$\eta^2 = \frac{SS_{effect}}{SS_{total}} \quad (22)$$

Cohen's (1992) recommended effect sizes are, then, applied for the interpretation of effect size measures (i.e., small: $\eta^2 \leq .06$, medium: $.06 < \eta^2 \leq .14$, and large: $\eta^2 > .15$) to focus the discussion on the factors that are most substantially related to parameter estimation accuracies and their inferences. The multi-way univariate ANOVAs and eta-analyses were computed using PROC GLM in SAS.

In addition to the multi-way univariate ANOVA analyses, graphical representations including box plots, histograms and line graphs for each outcome variable are presented across simulation conditions. Also, marginal means for each simulation condition factor were computed to summarize the results by simulation conditions. The computed marginal means are, then, illustrated with graphical representations.

CHAPTER FOUR: RESULTS

This section consists of the findings of the study. The chapter reports each dependent variable (bias, RMSE, CI coverage rates, CI widths, and statistical power/Type I errors) of the study. In addition, η^2 values from ANOVA analyses of each dependent variable provide an indication of the impact of each of the simulation design factors, non-normality factors and their interactions. The estimation accuracy (bias and RMSE) and statistical inference (CI coverage rate, CI width, statistical power, and Type I error) of the treatment effect parameter is sequentially reported across independent variables of the study by using box plots and bar graphs. Then, the estimation accuracy (bias and RMSE) for the parameters other than the treatment effect of the models are additionally provided. Finally, the convergence rates of the models are provided in the end of the chapter.

Bias for the Treatment Effect

The bias of the treatment effect was computed by taking an average of the bias across replications. Note that bias and relative bias are equivalent because the population value for the treatment effect was set for 1. The complete bias/relative bias results for the treatment effect are shown in Table 5 in Appendix A. In addition, to identify simulation factors that have a substantial effect on the bias, two univariate ANOVA analyses were conducted for the within- and between-series estimators, individually. The η^2 values from these ANOVA analyses are also presented in Table 24 in Appendix B.

Based on ANOVA analyses for the bias of the treatment effect, the interaction of the skewness and kurtosis factor (skewness*kurtosis) had a medium effect size (η^2 = .13) and the other simulation factors (e.g., *I**kurtosis, *J**kurtosis, *I*, skewness, *I**skewness, kurtosis, *I***J*, and *J**skewness) had small effect sizes (η^2 < .03) for the within-series model. For the between-series model, skewness*kurtosis had a medium effect (η^2 = .07) and the other factors (*I**kurtosis, *J**kurtosis, *I***J*, *I**skewness, *J**skewness, skewness, and kurtosis) had small effect sizes (η^2 < .06).

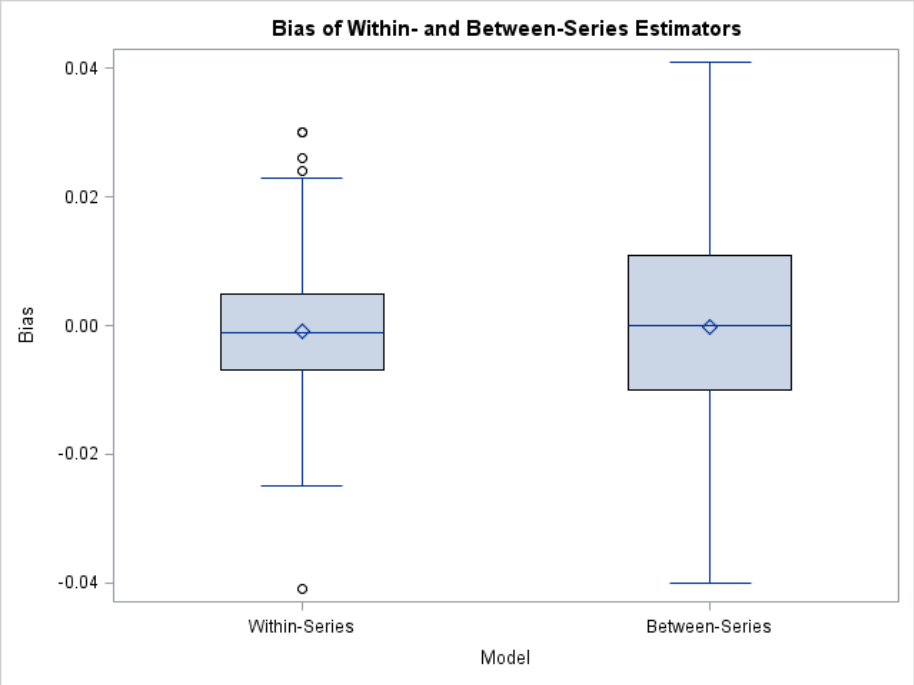


Figure 6. Box plots: Marginal bias of the within- and between-series estimators across simulation conditions

Figure 6 presents box plots for marginal bias of the within- and between estimators. As shown in Figure 6, the bias of the within- and between-series estimators for the treatment effect was distributed less than 5% bias of the population value across simulation conditions. The

marginal mean bias of the two estimators were close to zero, and they were ranged from -.04 and .04 across conditions. The box plots also indicated that the within-series estimator had ales variance in the bias distribution than the between-series estimator.

Figures 7 and 8 represent bias of the within- and between-series estimators across skewness and kurtosis, respectively. As shown in Figures 7 and 8, minimal bias was observed across various degrees of skewness and kurtosis. The bias of the treatment effect estimate for the within- and between-series models was ranged between -.04 and .04 and the minimal bias were consistently found as skewness and kurtosis increased. Although ANOVA analyses indicated the interaction between skewness and kurtosis explained 13% of the variance in bias values, the variance was so small that there was no meaningful effect of skewness and kurtosis on bias.

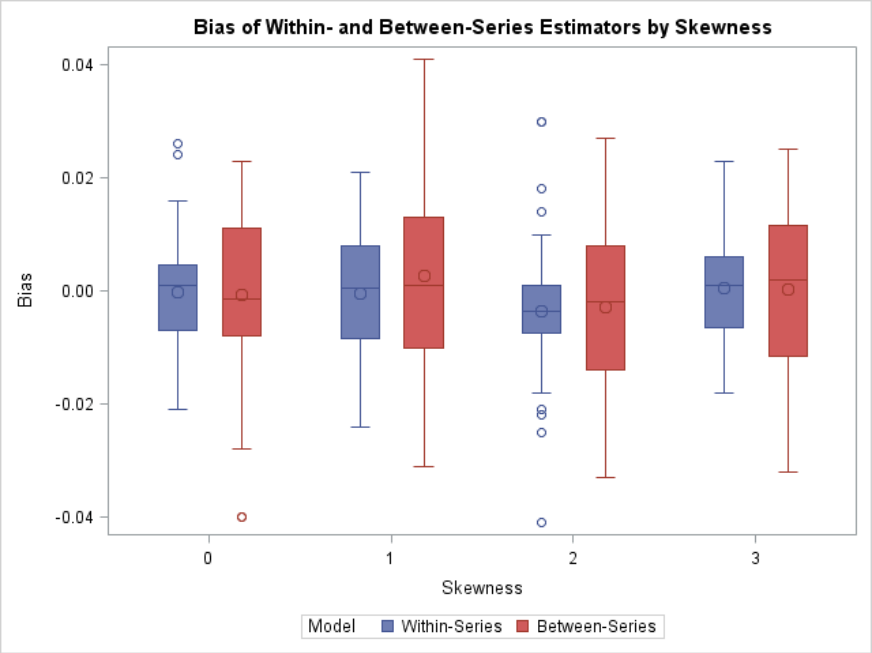


Figure 7. Box plots: Bias of the within- and between-series estimators across skewness

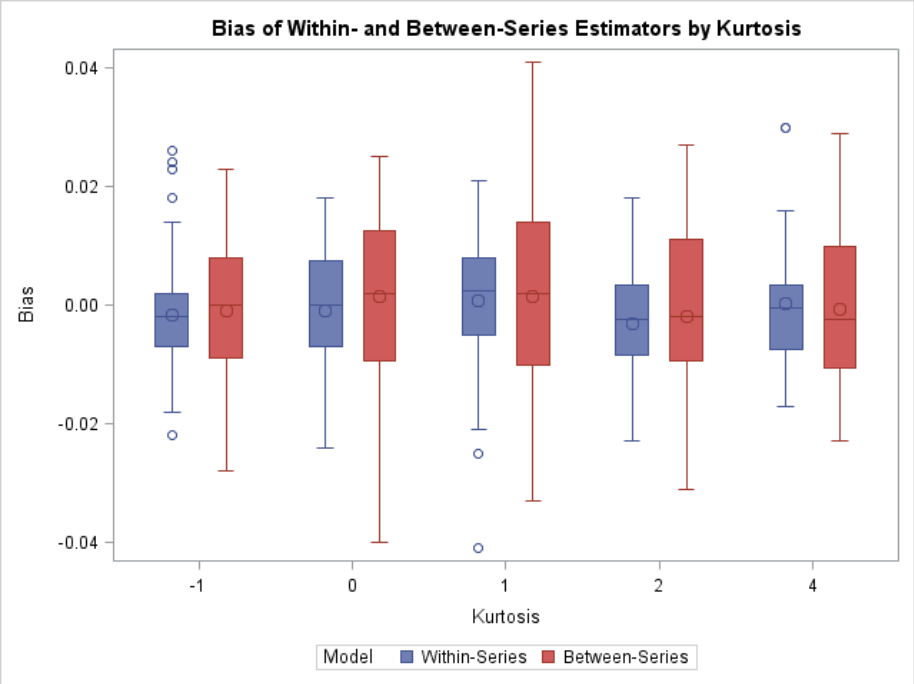


Figure 8. Box plots: Bias of the within- and between-series estimators across kurtosis

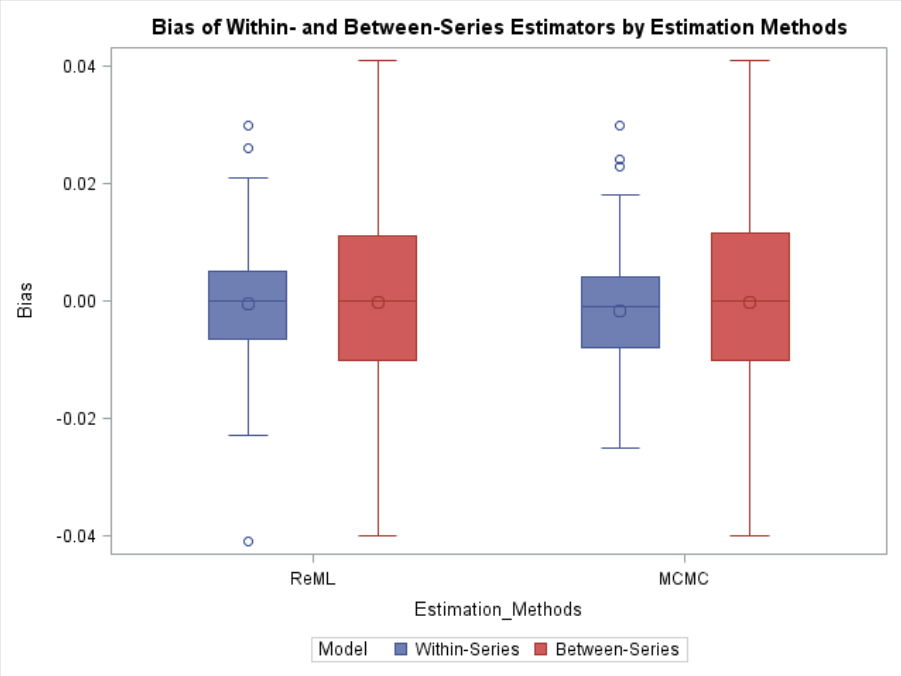


Figure 9. Box plots: Bias of the within- and between-series estimators across estimation methods

Figure 9 shows bias box plots of the within- and between-series estimators across estimation methods. Based on ANOVA analyses and Figure 9, it was found that ReML and Bayesian estimations for the treatment effect estimate were almost indistinguishable. No significant difference between two estimation methods was found based on the results of the study. Both estimation methods produced minimal bias across simulation conditions.

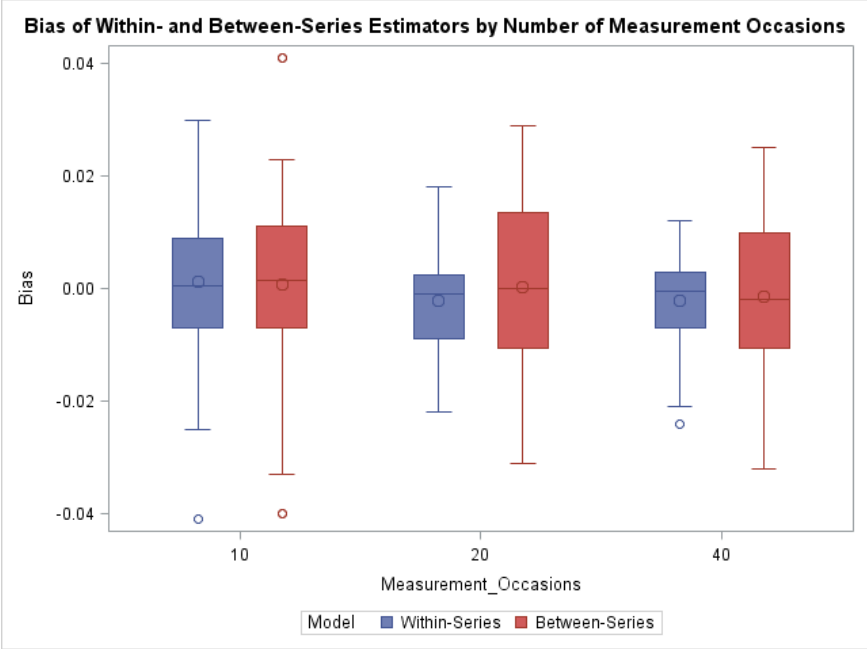


Figure 10. Box plots: Bias of the within- and between-series estimators across number of measurement occasions

Figures 10 and 11 represent bias of the within- and between-series estimators across the number of measurement occasions and the number of the participants, respectively. As shown in Figures 10 and 11, the marginal mean bias for the within- and between-series model was near zero across conditions. For the within-series model, standard deviation of the bias decreased as both the numbers of the measurement occasions and participants increased. For the between-

series model, however, the standard deviation of the bias only decreased as the number of participants increased. Marginal bias of the within- and between-series estimators was consistently near zero as numbers of measurement occasions and participants increased.

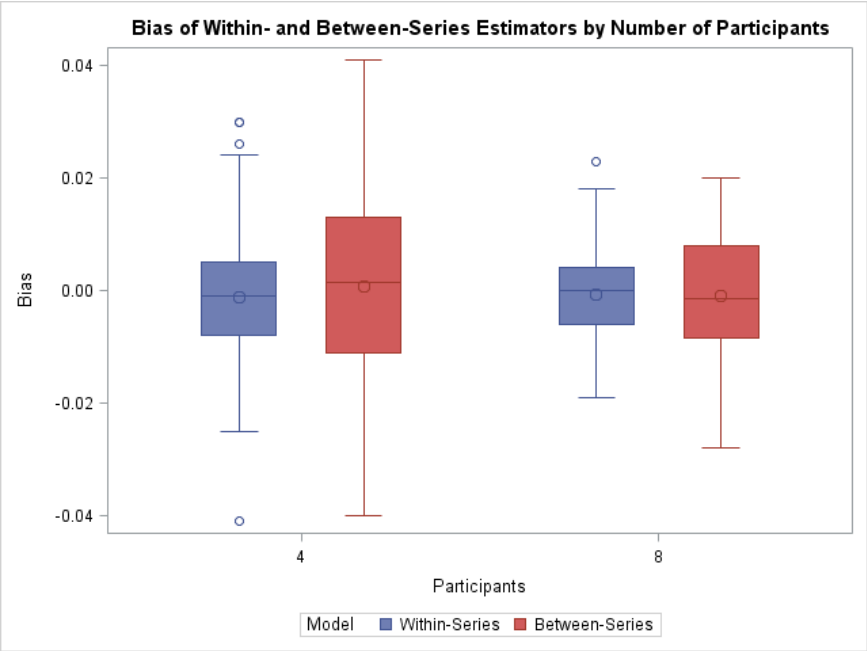


Figure 11. Box plots: Bias of the within- and between-series estimators across number of participants

RMSE for the Treatment Effect

RMSE for the treatment effect estimate was computed by taking a square root of the average squared bias across replications. Note that RMSE includes not only squared bias but also variance across replications. The variance across replications is also considered as sampling error of the simulation study.

In Table 6 of Appendix A, the complete RMSE results of the treatment effect across simulation conditions are presented. In addition, in Table 25 of Appendix B, two univariate ANOVA analyses for the within- and between-series models are presented to identify simulation

factors that have a significant effect on RMSE. Based on the ANOVA analyses, the number of measurement occasions had the largest effect size (η^2 = .62), followed by the number of participants (η^2 = .32) for the within-series model. Also, the interaction between the number of measurement occasions and participants (*I***J*) and estimation method had a small effect size on RMSE when the within-series model was used (η^2 < .03). For the between-series model, the number of participants had the largest effect size (η^2 = .93) and the number of measurement occasions had a medium effect size (η^2 = .07) on RMSE. It is noteworthy that neither skewness nor kurtosis of the level-1 error variance had a significant effect on RMSE for both within- and between-series models (η^2 = .00).

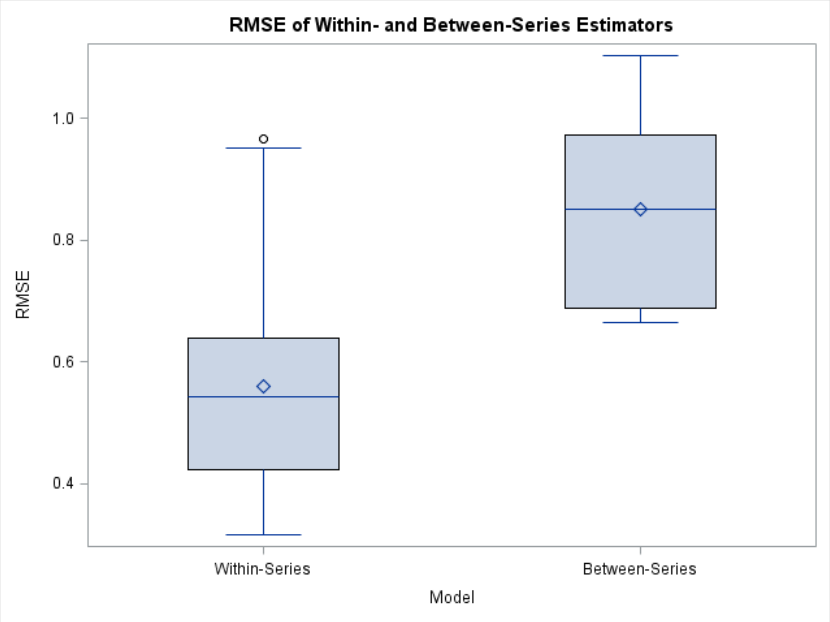


Figure 12. Box plots: Marginal RMSE of the within- and between-series estimators

Marginal RMSE box plots of the within- and between-series estimators for the treatment effect across simulation conditions are illustrated in Figure 12. The marginal box plots of RMSE

show that the within-series estimator produced less RMSE of the treatment effect estimate than the between-series estimator. The marginal mean RMSE was .65 for the within-series model as opposed to .85 for the between-series model. Also, the standard deviations of RMSE across simulation conditions were .18 and .15 for the within- and between-series model, respectively.

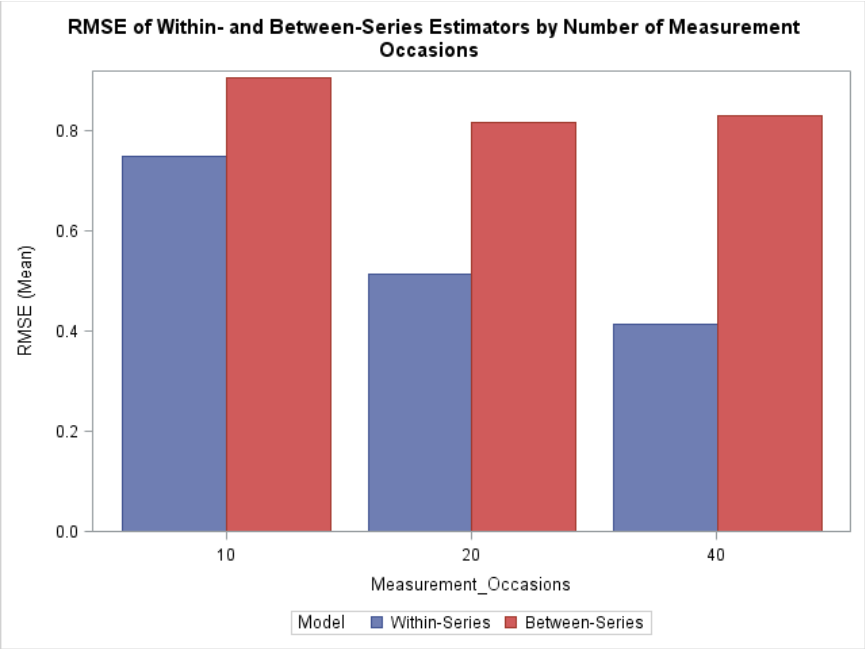


Figure 13. Bar graphs: RMSE of the within- and between-series estimators across number of measurement occasions

Based on ANOVA analyses on RMSE, bar graphs for both within- and between-series models were also created across the numbers of measurement occasions and participants. Figures 13 and 14 represent bar graphs of RMSE across the number of measurement occasions and the number of participants, respectively. As shown in Figure 13, RMSE for the within-series model decreased substantially as the number of measurement occasions increased. However, consistent with ANOVA analyses, RMSE for the between-series model decreased minimally as the number

of measurement occasions increased. From the bar graphs as illustrated in Figure 14, RMSE for both within- and between-series models decreased significantly as the number of participants increased.

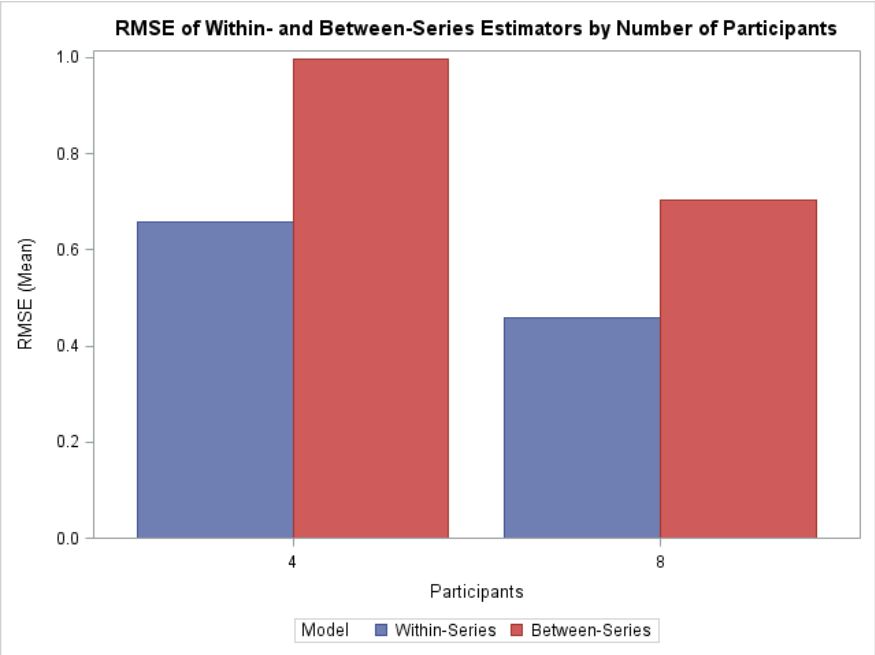


Figure 14. Bar graphs: RMSE of the within- and between-series estimators across number of participants

CI Coverage Rate for the Treatment Effect

To examine accuracy of the statistical inference for the within- and between-series models on the treatment effect, CI coverage rate was computed. Two estimation methods, ReML and Bayesian, computed the CIs differently: ReML used Kenward-Roger inference method computing small sample adjusted standard error and degrees of freedom, whereas Bayesian used highest posterior density (HPD) of the Markov chain samples to compute CI. CI coverage rate

was then obtained by computing the proportion of replications in which the population value was inside of the computed confidence interval. The nominal level for CI was set as .95.

The complete CI coverage rate table is shown in Table 7 in Appendix A. To identify simulation factors that have a significant effect on CI coverage rate, univariate ANOVA analyses for the within- and between-series models were conducted and presented in Table 26 in Appendix B. Based on ANOVA analyses, the number of measurement occasions had the largest effect size (η^2 = .41), followed by the estimation method (η^2 = .17) for the within-series model. The interaction between the number of participants and estimation method (*J**Est) had a medium effect size (η^2 = .07) and other factors (e.g., kurtosis, *I**Est, skewness*kurtosis, *I**skewness, *I***J*, *I**kurtosis, and *J*) had a small effect (η^2 < .02) when the within-series model was used. For the between-series model, a similar pattern was observed. The number of measurement occasions also had the largest effect size (η^2 = .40) and the estimation method had the second largest effect size (η^2 = .12). The other factors (e.g., *J*, *I**kurtosis, skewness*kurtosis, *I**skewness, *I***J*, kurtosis, *J**kurtosis, *J**Est, and *J**skewness) showed a small effect (η^2 < .04) for the between-series model. Similar to the previous results, skewness and kurtosis had a minimal effect on CI coverage rate for both within- and between-series models.

Figure 15 presents box plots of marginal CI coverage rate for the within- and between-series models across simulation conditions. As shown in Figure 15, the within-series model produced acceptable ranges of the treatment effect CI coverage rate across conditions. However, the between-series model yielded an under-coverage rate less than the acceptable CI coverage rate. Marginal mean CI coverage for the within-series model was .94 whereas that for the between-series model was .87. The standard deviation of CI coverage rate across simulation condition was .01 for both within- and between-series model.

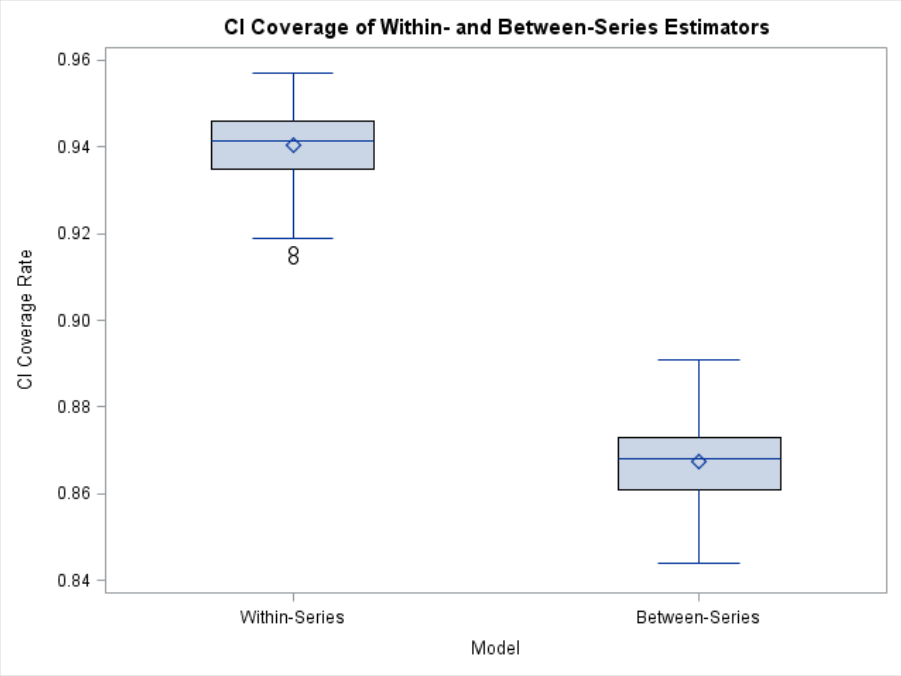


Figure 15. Box plots: Marginal CI coverage rate of the within- and between-series models

As ANOVA analyses indicated, the measurement occasion and estimation method had the large effect size of CI coverage rate, thus, the marginal box plots for each design factor were created. Figure 16 shows that CI coverage rate of the within- and between-series models across the number of measurement occasions. As can be seen in Figure 16, CI coverage rate decreased as the number of measurement occasions increased. The same pattern was observed for the between-series model. Across conditions, between-series consistently showed lower CI coverage rate than the within-series model. Marginal CI coverage rate of the within-series model was .95, .94 and .93 when the number of measurement occasions was 10, 20 and 40, respectively. Similarly, marginal CI coverage rate of the between-series model was .87, .87 and .86 across 10, 20 and 40 measurement occasions.

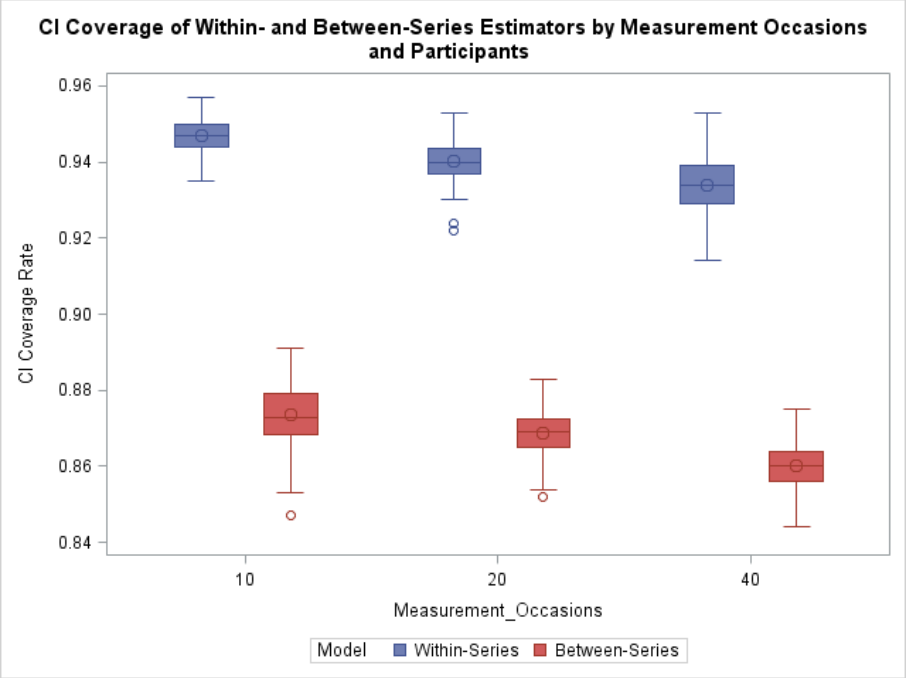


Figure 16. Box plots: CI coverage rate of the within-series model across numbers of measurement occasions

Figure 17 illustrates box plots of CI coverage rate for the within- and between-series model across estimation methods. Overall, ReML with Kenward-Roger and Bayesian estimation and inference methods showed comparable performances. CI coverage rate of both approaches produced closer values to the nominal level and their values across simulation conditions were distributed in the acceptable range. For example, the marginal mean CI coverage rate for the ReML and Bayesian methods were .94 when the within-series model was used. In addition, a similar pattern was found for the between-series model. Although the overall values were lower than a nominal level, ReML and Bayesian produced similar CI coverage rate across conditions. For example, the marginal mean CI coverage rate for ReML estimation was .87 and the corresponding value for the Bayesian estimation was .86.

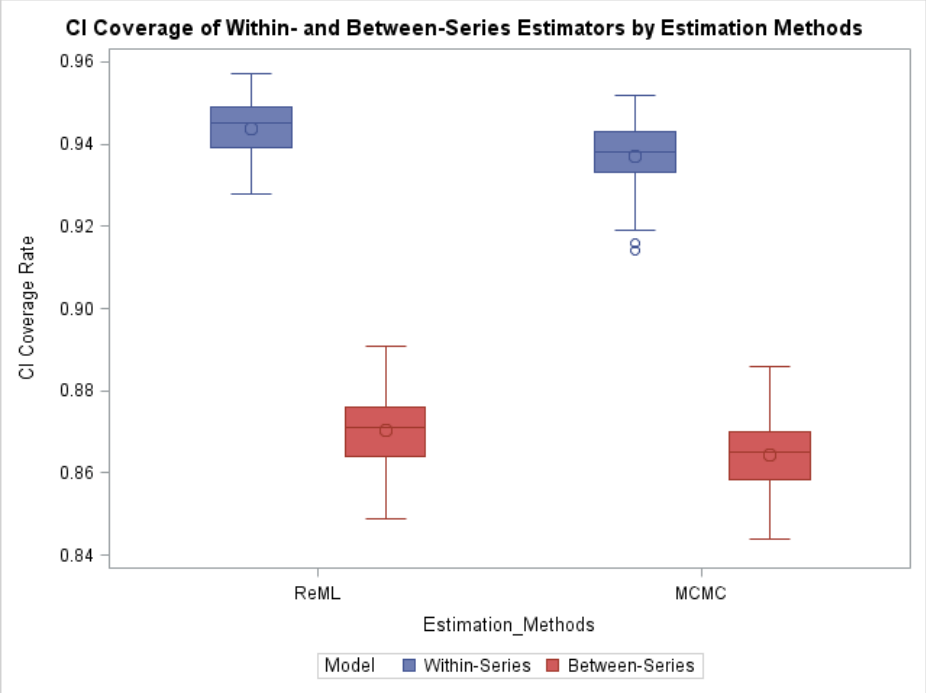


Figure 17. Box plots: CI coverage rate of ReML and MCMC for the within- and between-series models

CI Width for the Treatment Effect

In addition to CI coverage rate, CI width was computed as well across simulation conditions. The complete CI width table is shown in Table 8 in Appendix A. Similar to the previous outcome measures, univariate ANOVA analyses for the within- and between-series models were conducted separately and presented in Table 27 in Appendix B. Based on ANOVA analyses, the two largest effect size design factors on CI width were the number of measurement occasions and the number of participants for both models. The number of measurement occasions had the largest effect size (η^2 = .57), followed by the number of participants (η^2 = .17) for the within-series model. The interaction between the numbers of measurement occasions and participants (*I***J*) had a small effect size (η^2 = .02) when the within-series model was used. For

the between-series model, on the contrary, the number of participants had the largest effect size (η^2 = .88), followed by the number of measurement occasions (η^2 = .11). Similar to the previous results, skewness and kurtosis had a minimal effect on CI width for both within- and between-series models.

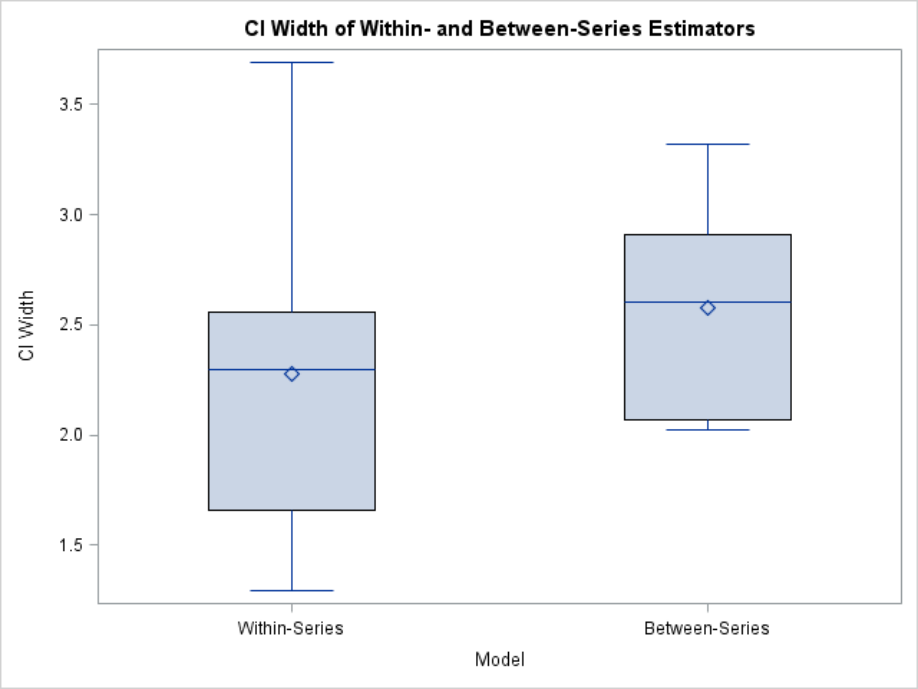


Figure 18. Box plots: Marginal CI width of within- and between-series estimators

Figure 18 shows two the marginal box plots of CI width for the within- and between-series models. As shown in Figure 18, no substantial difference between two models was found. For example, the marginal mean CI width for the within-series model was 2.27 and the corresponding value for the between-series model was 2.57 across conditions. Note that the within-series model, however, showed more variability of CI width across simulation conditions than the between-series model. For example, the standard deviation of CI width across

conditions was .75 for the within-series model whereas the corresponding value was .46 for the between-series model.

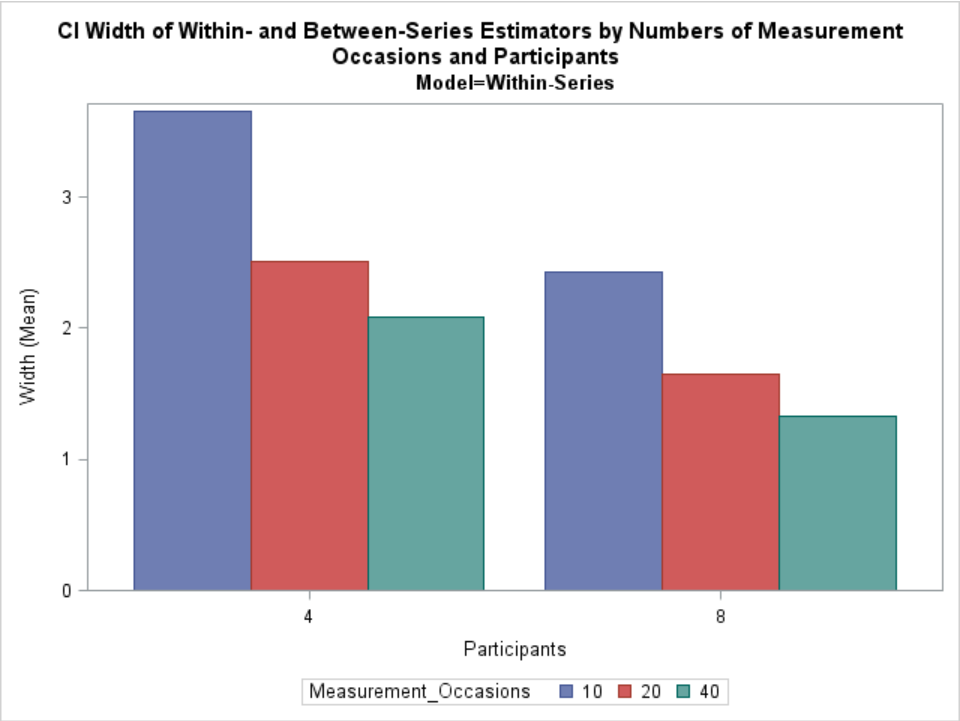


Figure 19. Bar graphs: CI width of the within-series estimator across numbers of measurement occasions and participants

Figures 19 and 20 represent two-way interaction CI width bar graphs across the number of measurement occasions and the number of participants for the within- and between-series models, respectively. As shown in ANOVA analyses on CI width, Figures 21 and 22 show that the numbers of measurement occasions and participants had the largest effect size for both within- and between-series models. For the within-series model, as the numbers of measurement occasions and participants increased, CI width decreased substantially. A minimum CI width was observed when the number of participants was eight and the number of measurement occasions

was 40. However, for the between-series model, the substantial decrement of CI width was found as the number of participants increased. Although the decreasing pattern of CI width was observed as the number of measurement occasions increased, the size of the effect was smaller than the number of participants.

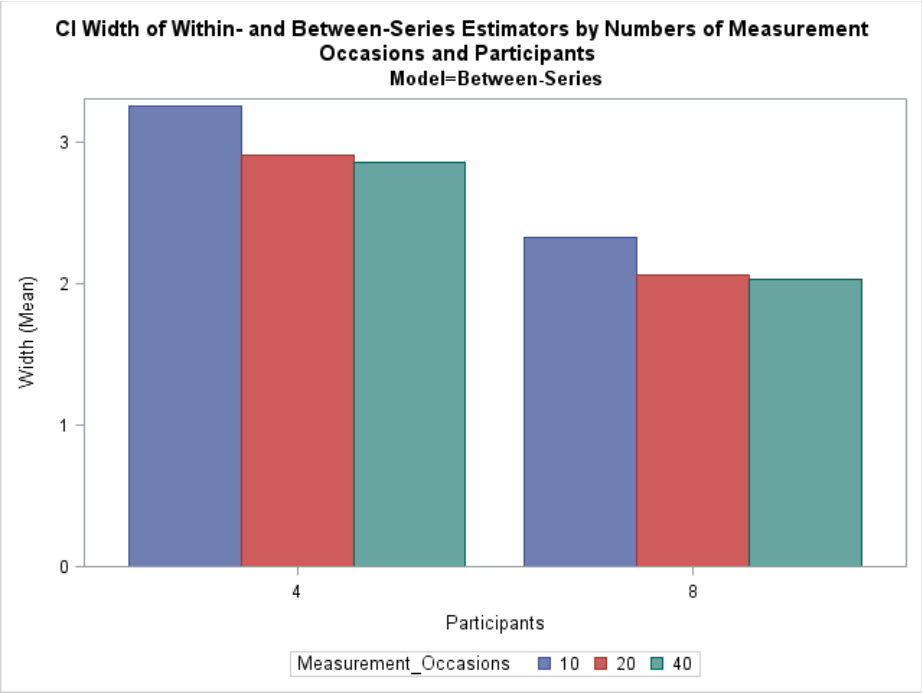


Figure 20. Box plots: CI width of the between-series estimator across numbers of measurement occasions and participants

Power/Type I Error for the Treatment Effect

Statistical power and Type I error of the treatment effect estimate were also dependent variables of the study. Tables 9 and 10 in Appendix A provide the complete power and Type I error tables across simulation conditions. Two-way ANOVA analyses on power were also conducted and the resulting η^2 values are provided in Table 28 in Appendix B. For the power condition, similar to RMSE and CI width, the numbers of measurement occasions and

participants had the largest effect sizes when the within-series model was used to estimate the treatment effect. The number of measurement occasions had the largest effect size (η^2 = .55), followed by the number of participants (η^2 = .41). The interaction between these two variables (*I***J*) had a minimal effect size (η^2 = .03). When the between-series model was used, the number of participants had the largest effect size (η^2 = .87), followed by the number of measurement occasions (η^2 = .11). Similar to the previous results, skewness and kurtosis had a minimal effect on the power of the test for the treatment effect estimate for both within- and between-series models.

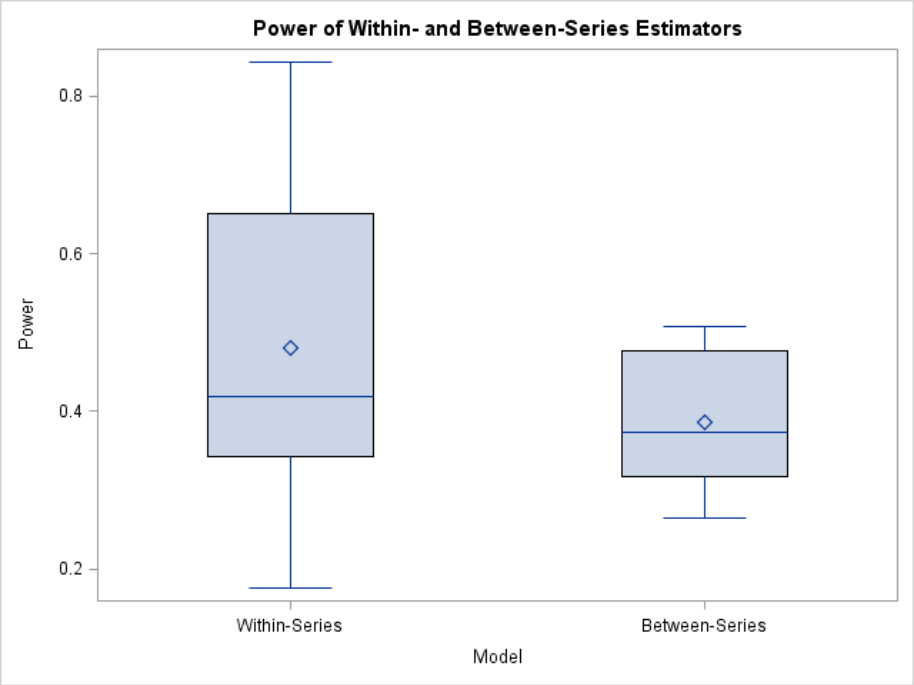


Figure 21. Box plots: Marginal statistical power of the within- and between-series estimators

Figure 21 shows the marginal mean power box plots for the within- and between-series models. It is prominent that the within-series model has higher power to detect the treatment effect than the between-series model. The marginal power for the within-series model was .48,

whereas the corresponding value for the between-series model was .38. In addition, the within-series model showed more variation than the between-series model.

As ANOVA analyses indicated, the numbers of measurement occasions and participants had the largest effect sizes on statistical power. Thus, Figures 22 and 23 illustrated interaction bar graphs across the number of measurement occasions and the number of participants for the within- and between-series models, respectively. Consistent with expectation, as the number of measurement occasions and participants increased, statistical power increased substantially with the within-series model. On the other hand, statistical power increased considerably as the number of participants increased only for the between-series model.

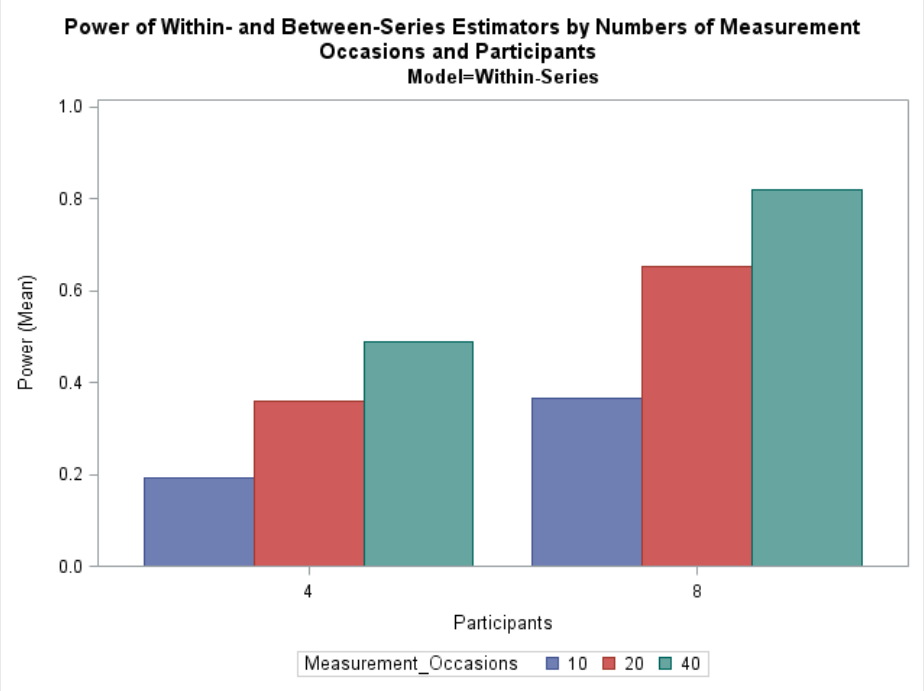


Figure 22. Bar graphs: Statistical power of the within-series estimator across numbers of measurement occasions and participants

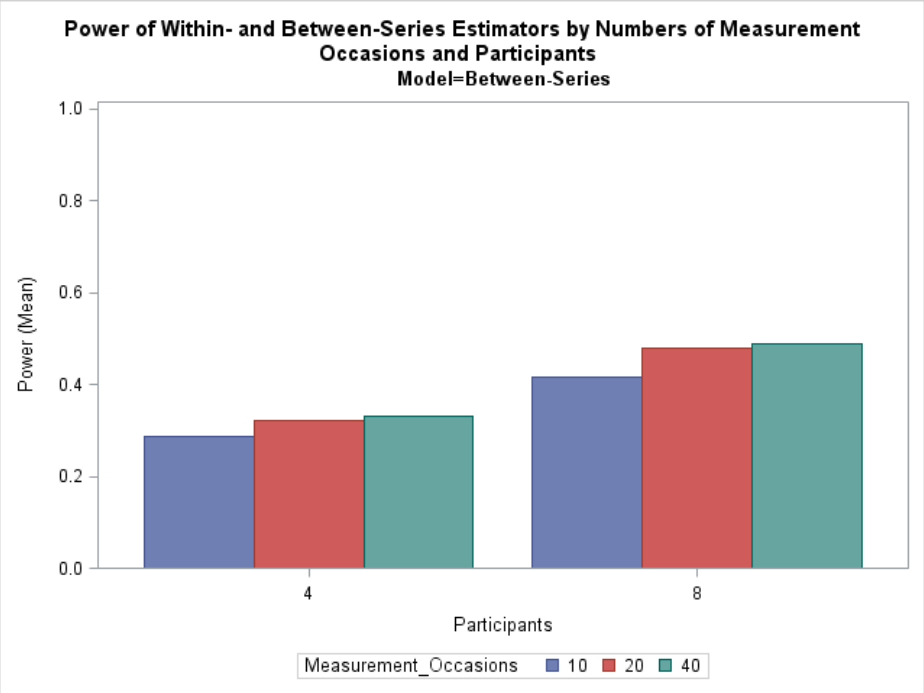


Figure 23. Bar graphs: Statistical power of the between-series estimator across numbers of measurement occasions and participants

For the Type I error result, Figure 24 shows box plots of the marginal mean Type I error results between two models. Type I error was well-controlled for the within-series model, whereas inflated values were consistently found for the between-series model. The marginal Type I error for the within-series model was .06, as opposed to .13 for the between-series model across simulation conditions. For both within- and between-series models, maximum Type I errors, .09, and .16 were observed when numbers of participants and measurement occasions were 8 and 40, respectively and MCMC was used. For the skewness and kurtosis conditions, consistent pattern of the Type I error results was found.

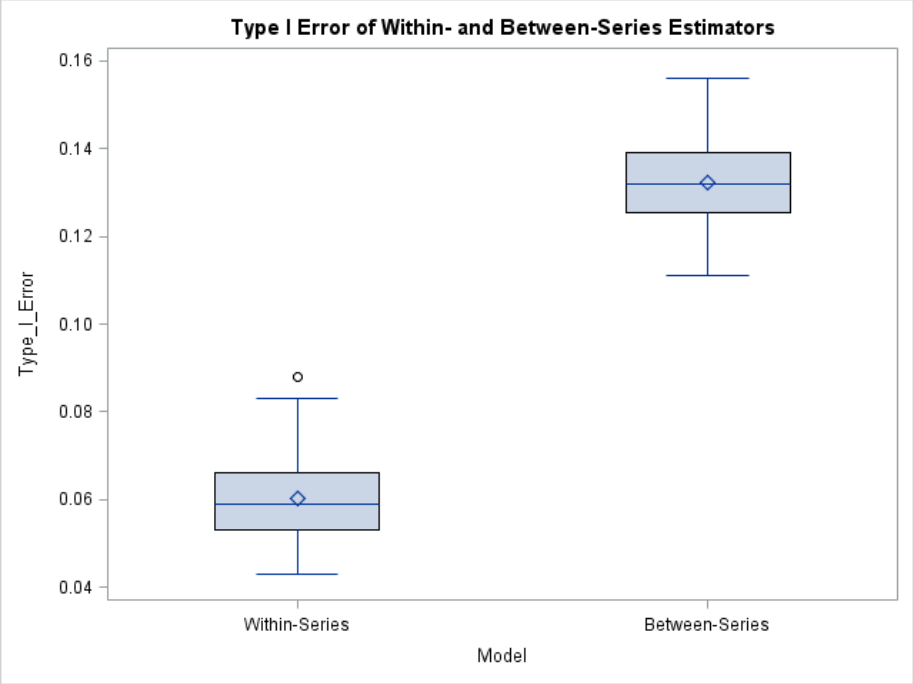


Figure 24. Box plots: Marginal Type I error of the within- and between-series estimators

Bias and RMSE for the Other Fixed Effects

Bias and RMSE of the fixed effect parameters other than the treatment effect for the within-series model were computed in the current study. Note that the within-series model was specified with four parameters including intercept, treatment effect, time effect for the baseline phase, and change in time effect with treatment. Because the population values for the other fixed parameters were set as zero, bias of the parameter estimates was computed rather than relative bias.

The complete tables of bias and RMSE are presented in Tables 12 – 17 in Appendix A. Figures 25 and 26 also represent bias and RMSE box plots of intercept, time effect for the baseline, and change in time effect with treatment, respectively. Overall, minimal bias was observed for the fixed effect parameters across simulation conditions. As shown in Figure 25, intercept, baseline time effect, and change in time effect with treatment parameters had bias

values that ranged from -.02 to .02 and marginal bias values that were close to zero. In addition, RMSE distributions of fixed effect parameters of the within-series model are shown in Figure 26. Box plots of RMSE indicate intercept parameters had larger RMSE than the other fixed effect parameters. Marginal RMSE of the intercept, baseline time effect and change in time trend with treatment parameters were .38, .06, and .10, respectively.

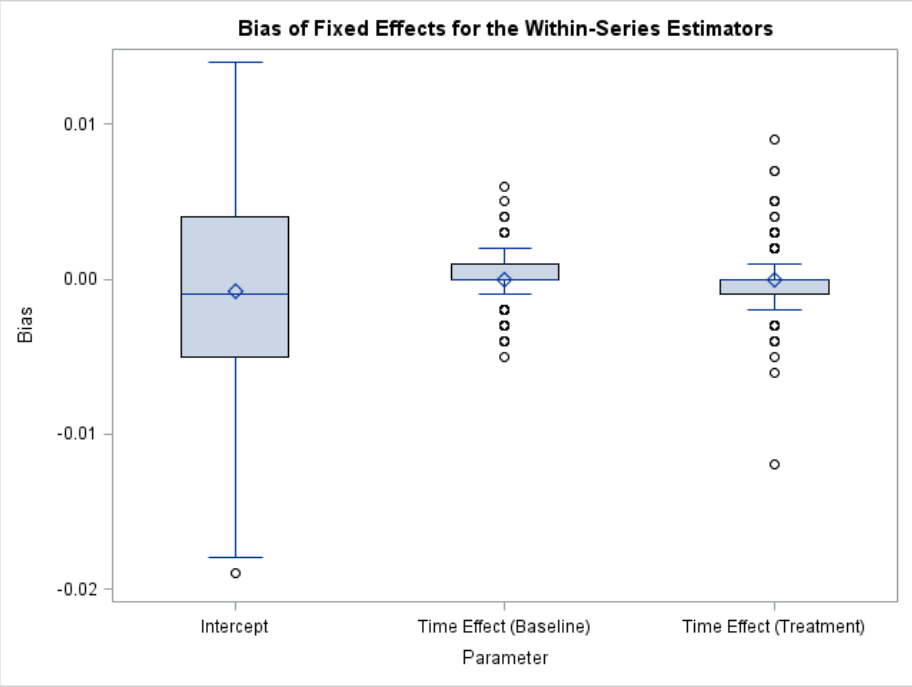


Figure 25. Box plots: Bias of the other fixed effect parameters for the within-series model

As shown in previous results, no substantial difference between ReML and Bayesian estimation methods was found in bias and RMSE of the fixed effect parameters. As expected, RMSE for the other fixed effect parameters decreased considerably as the numbers of measurement occasions and participants increased. Minimal RMSE were found when the measurement occasion was 40 and the number of participants was eight. Consistent with

previous results, skewness and kurtosis of the level-1 error variance in the within-series model had no significant effect on bias and RMSE of the other fixed effect parameters.

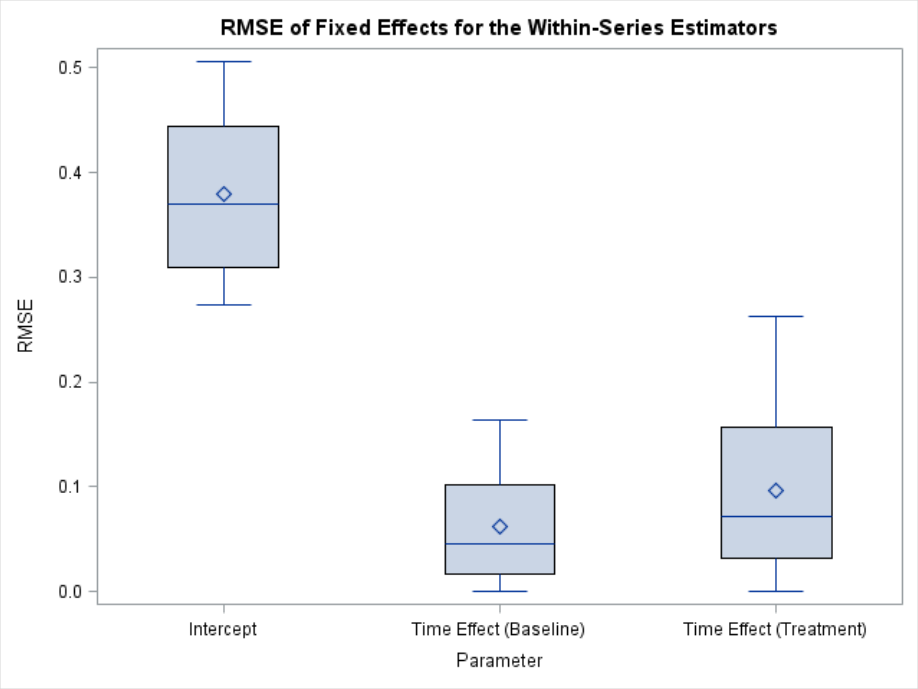


Figure 26. Box plots: RMSE of the other fixed effect parameters for the within-series model

Bias and RMSE for the Level-1 Error Variance

Bias and RMSE of the leve-1 error variance for the within- and between-series models were computed. The complete results are presented in Tables 18 and 19 in Appendix A. Note that the level-1 error variance was standardized after the Fleishman's transformation. Thus, population mean and variance for the level-1 error variance remained zero and one, respectively. Also, since the level-1 error variance for the between-series model contains the within- and between-participant variations, the population value was set combining level-1 error variance and level-2 error variance ($\sigma^2 + \tau_{00}^2$). This parameter value is appropriate for the baseline phase, but in the treatment phase the variance would be larger because of the variation in the treatment

effect ($\sigma^2 + \tau_{00}^2 + \tau_{11}^2$). As a consequence, one would anticipate the estimated variance values from the between-series model will exceed the baseline variance parameter value. A follow-up study at the end of this chapter presents results for a between-series model with separate variance estimates for the baseline and treatment phases.

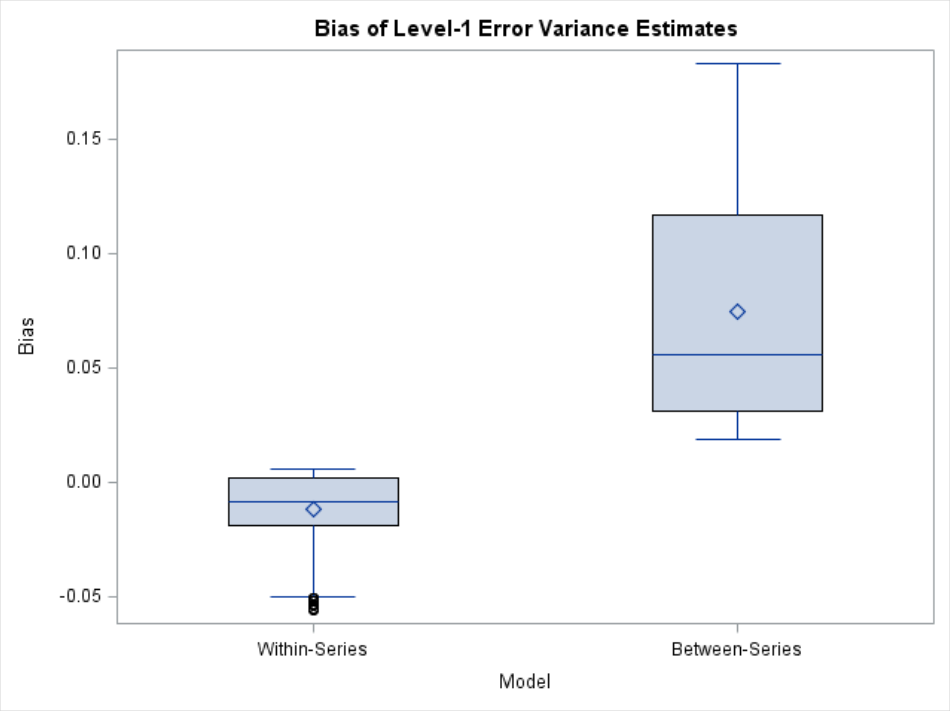


Figure 27. Box plots: Bias of the level-1 error variance

Figures 27 and 28 illustrate box plots of bias and RMSE of level-1 error variance for the within- and between-series models. As shown in Figure 27, the within-series model estimated the level-1 error variance with relatively smaller bias than the between-series model across simulation conditions. The bias of the level-1 error variance was ranged from -.06 to .01. Also, the RMSE values were ranged from .00 to .12 across conditions. However, for the between-series model, bias and RMSE ranged from .02 to .18 and from .04 to .25, respectively, across

conditions. Maximum bias was found when the number of measurement occasions was 10 and the number of participants was four.

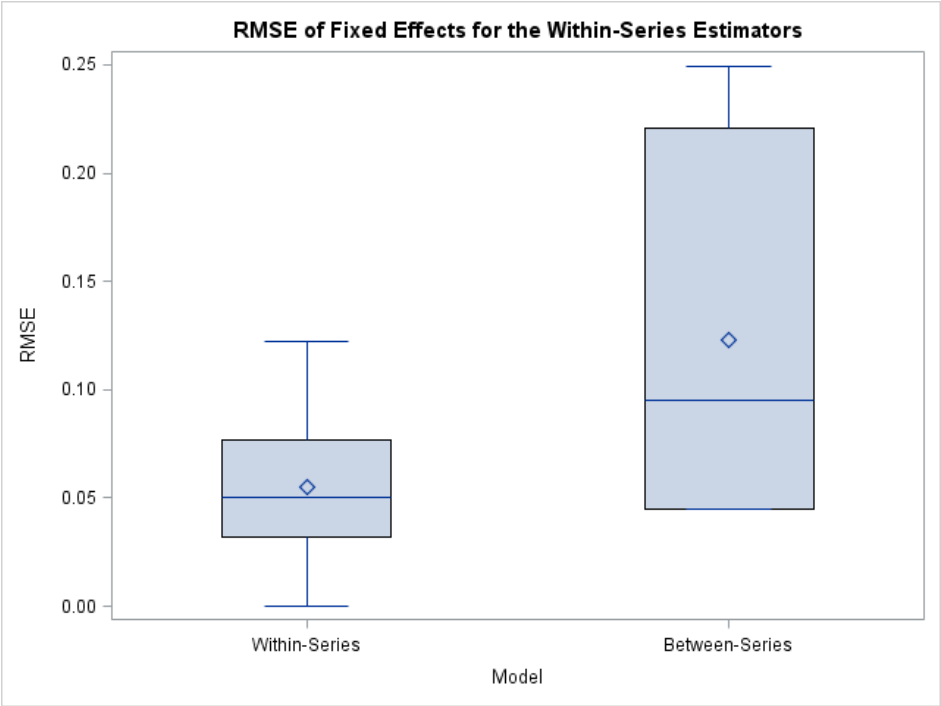


Figure 28. Box plots: RMSE of the level-1 error variance

Consistent with the previous results, skewness and kurtosis of the level-1 error variance did not affect the bias and RMSE results. For both within- and between-series models, bias, and RSME of the level-1 error variance were similar across skewness and kurtosis conditions. In addition, ReML and Bayesian estimations did not show a substantial difference. As expected, RMSE decreased as the numbers of participants and measurement occasions increased.

Bias and RMSE for the Level-2 Error Variance

Bias and RMSE of the level-2 error variance estimate were computed for the within-series model only. The complete tables are presented in Tables 20 – 23 in Appendix A. Note that the population value was 0.5 for the intercept (τ_{00}^2), and treatment effect (τ_{11}^2) level-2 error variances, and 0 for the other parameter level-2 error variances, respectively. Bias and RMSE were computed only for τ_{00}^2, and τ_{11}^2 because the population values were non-zero and relative bias for the corresponding parameters can be obtained.

Consistent with expectation, level-2 error variance estimates of the within-series model were biased across conditions. Relative bias were ranged from -.27 to .12, and from -.17 to .09 for τ_{00}^2 and τ_{11}^2, respectively. Also, RMSE values ranged from .18 to .50 for τ_{00}^2 and from .12 to .61 for τ_{11}^2. Interestingly, it was found that Bayesian estimation showed better accuracy estimating the level-2 error variance than ReML. Figures 29 and 30 illustrate bias and RMSE box plots for τ_{00}^2 and τ_{11}^2 across ReML and Bayesian estimations. Overall, the marginal mean bias of Bayesian estimation was close to zero for both τ_{00}^2 and τ_{11}^2, whereas, the corresponding values of ReML were near -.10 and -.15 for τ_{00}^2 and τ_{11}^2. In addition, as shown in Figure 30, RMSE of τ_{00}^2 and τ_{11}^2 estimates for Bayesian estimation were considerably smaller than ReML estimation. This finding was consistently found across simulation conditions.

The skewness and kurtosis of the level-1 error variance had a minimal effect on the accuracy of level-2 error variance estimation. Also, estimation accuracy increased as the numbers of the measurement occasions and participants increased.

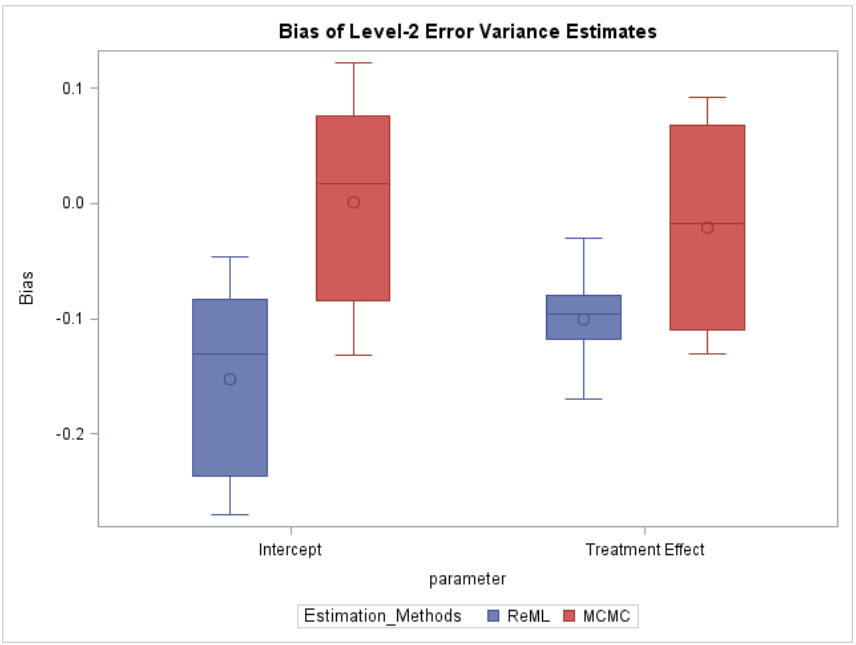


Figure 29. Box plots: Bias of the level-2 error variance for the within-series model across estimation methods

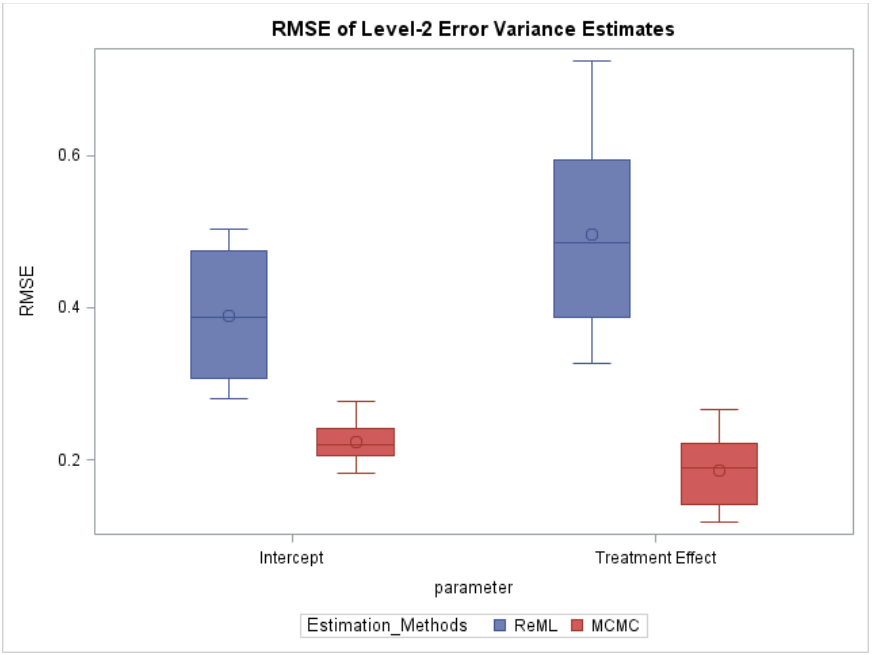


Figure 30. Box plots: RMSE of the level-2 error variance for the within-series model across estimation methods

Convergence Rate

The convergence rates for both within- and between-series models were computed across replications. Also, ReML and Bayesian estimation convergence rates were computed as well. Results showed that both within- and between-series models were 100% converged across replications and simulation conditions when ReML estimation was used. However, when Bayesian estimation method was used, average 96% convergence rate was observed across simulation conditions. Note that, in the current study, convergence of Bayesian estimation was determined based on Geweke' statistics.

Follow-Up Study

For the between-series model, although minimal bias was produced across simulation conditions, statistical inferences including CI coverage rate, and Type I error were problematic. For example, it was found that the between-series model yielded substantially lower CI coverage rate than the nominal level for the treatment effect estimate. In addition, Type I error was substantially inflated across simulation conditions. This finding was consistent when the level-1 error variance was normally distributed. Note that variance structure for the between-series model in the current study was specified for homogeneous across the baseline and treatment phases. However, the data generation model included the level-2 error variances for intercept and treatment effect and fitting the data with the homogeneous variance model represents a form of model misspecification. Based on the previous study (Ferron et al., 2014), the heterogeneous variance structure can be specified for the between-series model and it seems worthwhile to explore whether the between-series model would perform differently across non-normality

conditions if the heterogeneous variance is specified. Thus, the follow-up study was conducted to examine the performance of the between-series model with heterogeneous variance structure.

In the follow-up study, the same skewness (0, 1, 2, and 3) and kurtosis (-1, 0, 1, 2, and 4) of the level-1 error variance conditions were considered to examine the effect of the non-normality. Also, two sample size conditions (small [I = 10, J = 4], and large [I = 40, J = 8]) were considered. Because ReML and Bayesian methods showed indistinguishable performance in the earlier study, only ReML was used to estimate the heterogeneous variance between-series model in the follow-up study. The data generation model and population values were the same as in the earlier study. For the dependent variable of the follow-up study, bias, RMSE, CI coverage rate, CI width, power and Type I error of the treatment effect were computed. The number of the replications was set as 3000 to keep the consistency.

Tables 3 and 4 show the comparisons of the homogeneous and heterogeneous variance between-series models. Table 3 shows accuracy of the treatment effect estimation and inference for the small sample size conditions, and Table 4 shows the corresponding values for the large sample size conditions. Consistent with the homogeneous variance model, the heterogeneous variance model produced the minimal bias across conditions. Also, RMSE of the treatment effect estimate for the heterogeneous variance was comparable to the homogeneous variance model results. Similar to the previous results, skewness and kurtosis of the level-1 error variance did not have a significant effect. No substantial difference or pattern was observed as skewness and kurtosis increased. However, as shown in both Tables 3 and 4, CI coverage rate was in the acceptable ranges across simulation conditions for the heterogeneous variance model. In addition, relatively wider CI width and lower statistical power were observed. Lastly, Type I

error was well-controlled for the between-series model across simulation conditions when heterogeneous variance structure was specified.

Table 3.
Homogeneous and Heterogeneous Variance Between-Series Models for the Small Sample Size Condition (I = 20 & J = 4)

		Homogeneous						Heterogeneous					
Skew	Kurt	Bias	RMSE	Cov	Wid	Pwr	TI	Bias	RMSE	Cov	Wid	Pwr	TI
0	-1	.023	1.10	**.867**	3.31	.299	.120	-.020	1.07	.947	5.52	.122	.054
	0	-.040	1.07	**.869**	3.31	.268	.128	.010	1.06	.943	5.53	.134	.063
	1	-.003	1.05	**.878**	3.31	.275	.120	-.020	1.07	.952	5.53	.127	.050
	2	.011	1.08	**.875**	3.32	.285	.112	.016	1.06	**.941**	5.48	.128	.049
	4	.000	1.07	**.869**	3.30	.284	.123	.031	1.08	.949	5.70	.126	.057
1	-1	-.004	1.10	**.861**	3.30	.285	.118	-.020	1.06	**.941**	5.52	.128	.054
	0	.000	1.05	**.878**	3.29	.284	.123	.001	1.07	.950	5.73	.128	.059
	1	.041	1.06	**.878**	3.30	.290	.124	.011	1.05	.944	5.32	.149	.059
	2	-.031	1.06	**.881**	3.31	.268	.123	-.025	1.09	.947	5.72	.130	.057
	4	.011	1.04	**.883**	3.30	.277	.113	.005	1.03	.948	5.34	.135	.055
2	-1	.002	1.06	**.872**	3.32	.270	.122	-.016	1.05	.947	5.71	.111	.062
	0	-.012	1.06	**.873**	3.31	.274	.129	.010	1.09	.943	5.51	.143	.059
	1	-.033	1.07	**.872**	3.30	.276	.132	-.018	1.05	.950	5.49	.111	.053
	2	-.006	1.07	**.879**	3.31	.288	.117	.000	1.05	.945	5.60	.149	.062
	4	-.011	1.06	**.872**	3.30	.279	.128	.023	1.05	**.939**	5.48	.153	.055
3	-1	.010	1.09	**.863**	3.31	.282	.129	-.012	1.04	.948	5.50	.125	.056
	0	.022	1.05	**.876**	3.30	.280	.126	-.004	1.04	.944	5.43	.111	.047
	1	-.009	1.03	**.889**	3.30	.265	.130	.023	1.06	.943	5.46	.153	.053
	2	.001	1.05	**.886**	3.31	.280	.120	.034	1.06	.947	5.62	.116	.055
	4	-.015	1.05	**.880**	3.31	.267	.128	-.002	1.02	.949	5.41	.134	.056

Note. I = number of measurement occasions, *J* = number of participants, Skew = skewness, Kurt = kurtosis, RMSE = root mean square error, Cov = CI coverage rate, Wid = CI width, Pwr = power, TI = Type I error. Values in Bold are not in the acceptable CI coverage range (.942 and .958).

Table 4.

Homogeneous and Heterogeneous Variance Between-Series Models for the Large Sample Size Condition (I = 40 & J = 8)

		Homogeneous						Heterogeneous					
Skew	Kurt	Bias	RMSE	Cov	Wid	Pwr	TI	Bias	RMSE	Cov	Wid	Pwr	TI
0	-1	-.028	.697	**.858**	2.04	.487	.137	.000	.684	.945	3.09	.255	.064
	0	.017	.686	**.860**	2.04	.508	.149	-.001	.679	.950	3.11	.251	.048
	1	-.009	.672	**.875**	2.04	.485	.136	-.004	.683	.946	3.12	.254	.048
	2	-.008	.683	**.860**	2.04	.491	.129	-.009	.681	.949	3.12	.251	.049
	4	-.002	.679	**.870**	2.04	.490	.128	.013	.696	.942	3.07	.281	.051
1	-1	.004	.704	**.856**	2.04	.493	.129	.016	.695	.957	3.03	.307	.045
	0	-.005	.683	**.865**	2.04	.480	.128	.012	.692	.946	3.12	.247	.051
	1	.020	.689	**.869**	2.04	.498	.136	-.016	.701	.944	3.08	.237	.056
	2	-.002	.675	**.868**	2.04	.484	.141	-.008	.680	.949	3.12	.241	.056
	4	-.010	.688	**.859**	2.04	.472	.136	.012	.688	.948	3.12	.257	.046
2	-1	.007	.688	**.859**	2.04	.482	.139	.007	.647	.954	3.13	.244	.057
	0	-.006	.691	**.854**	2.04	.469	.135	-.019	.701	.944	3.08	.233	.050
	1	.013	.684	**.863**	2.04	.500	.131	-.003	.694	.950	3.09	.311	.050
	2	-.014	.686	**.864**	2.04	.481	.139	-.013	.675	.948	3.08	.244	.046
	4	-.005	.667	**.870**	2.04	.486	.150	.007	.681	.946	3.13	.231	.051
3	-1	-.002	.668	**.874**	2.04	.493	.145	.007	.690	.950	3.10	.263	.046
	0	.003	.696	**.858**	2.04	.500	.142	.004	.693	.955	3.11	.302	.043
	1	-.015	.691	**.861**	2.04	.482	.129	.013	.687	.946	3.10	.248	.060
	2	.004	.701	**.859**	2.04	.496	.144	.009	.698	.948	3.13	.270	.062
	4	-.019	.683	**.864**	2.04	.482	.136	-.011	.683	.951	3.11	.236	.052

Note. I = number of measurement occasions, *J* = number of participants, Skew = skewness, Kurt = kurtosis, RMSE = root mean square error, Cov = CI coverage rate, Wid = CI width, Pwr = power, TI = Type I error. Values in Bold are not in the acceptable CI coverage range (.942 and .958).

CHAPTER FIVE: DISCUSSION

This chapter consists of summary of the study, findings, implications and applications for the applied single-case researchers and methodologists. Then, limitation and future research are further discussed.

Summary

The purpose of the study was to investigate the robustness of the within- and between-series estimators for the non-normal MB studies. The parameters of the within- and between-series models were estimated using two estimation methods: ReML and Bayesian. A Monte Carlo study was conducted under the conditions where level-1 error variances were generated from non-normal distributions manipulating skewness and kurtosis of the residuals' distribution. Fleishman's (1978) power transformation method was used to manipulate skewness and kurtosis of the distribution. Four modeling approaches were compared in the current study: a) two-level within-series model with ReML estimation and Kenward-Roger inference (Model 1), b) two-level within-series model with Bayesian estimation and inference (Model 2), c) between-series model with ReML estimation and Kenward-Roger inference (Model 3) and d) between-series model with Bayesian estimation and inference (Model 4).

The accuracy of parameter estimation and the statistical inference was systematically analyzed. Primarily, estimation accuracy and statistical inference for the treatment effect parameter of the fitted models were examined. Bias, relative bias, RMSE, CI coverage rates, CI

widths and statistical power/Type I error were computed as a function of specific design factors (number of measurement occasions and participants) and non-normality factors (amount of skewness, and kurtosis of the distribution).

A Monte Carlo study was designed to empirically evaluate the issues of violation of the normality assumption under various conditions. Data were generated varying the numbers of measurement occasions and participants, skewness and kurtosis of level-1 errors, and treatment effect sizes. Three levels of number of measurement occasions (10, 20, and 40), two levels of number of participants (4, and 8), four levels of skewness of the level-1 errors (0, 1, 2, and 3), five levels of kurtosis of the level-1 errors (-1, 0, 1, 2, and 4) and two levels of treatment effect sizes (0, and 1) were included. The conditions were chosen based on a preliminary survey of the published MB data and previous simulation studies. The analysis factors included four levels of modeling approaches (Models 1 - 4). Crossing all the data generation factors resulted in a total of 3 (number of measurement occasions) x 2 (number of participants) x 4 (level-1 error skewness) x 5 (level-1 error kurtosis) x 2 (treatment effects) = 240 simulation conditions. The simulation study results were analyzed by computing bias, relative bias, RMSE, CI coverage rate, CI width and the statistical power/Type I error for the treatment effect parameter and bias and RMSE for the other fixed effect and variance component parameters in the models.

Findings

Bias and RMSE of the Treatment Effect Estimate

The results of the study indicated that both within- and between-series models are robust to the non-normality of the level-1 error variance structure. The bias of the treatment effect estimate was consistently close to zero across various degrees of skewness and kurtosis. Relative

bias also was less than 5% of the population parameter value across simulation conditions. Similarly, RMSE of the treatment effect for the within- and between-series models was not affected by the non-normality of the level-1 errors. RMSE of the within- and between-series models, however, showed significant differences across conditions. The within-series model continuously yielded smaller RMSE of the treatment effect estimate than the between-series model. This finding is consistent with the previous study that compared the within- and between-series estimators (e.g., Ferron et al., 2014).

In addition, ReML and Bayesian estimation methods were compared in the current study. Based on the results of the study, no significant difference between two methods was found. Bias and RMSE of the treatment effect estimates were indistinguishable across simulation conditions. This finding can also be found in the previous research that compared ReML and Bayesian estimations (Moeyaert et al., 2016). Note that Moeyaert et al. (2016) compared ReML and Bayesian estimations using the two-level within-series model for MB data. Similar findings have also been found when ReML and Bayesian estimates of the fixed effects of multilevel models used with group designs have been compared for conditions where partially clustered data (Baldwin & Fellingham, 2013) and dichotomous outcome (Browne & Draper, 2006) was used.

Consistent with expectations and previous research that has examined the within-series model for single-case data structures with two levels (Ferron et al., 2009, 2010) and three levels (Moeyaert et al., 2013a, 2013b, 2014), treatment effect for both within- and between-series models were more precisely estimated as the number of measurement occasions and participants increased. More specifically, treatment effect of the within-series model was estimated more accurately as both the numbers of measurement occasions and participants increased. Treatment

effect of the between-series model, whereas, was more accurately estimated as only the number of participants increased. This finding is consistent with the previous study (Ferron et al., 2014).

CI Coverage Rate and CI Width of the Treatment Effect Estimate

CI coverage rate and CI width for the treatment effect were evaluated as a function of the skewness and kurtosis in level-1 error variance structure. Overall, the within-series model consistently showed acceptable CI coverage rate for both normal and non-normality conditions. This implies that CI coverage rate is not affected by non-normality in level-1 error variance structure and the within-series model is robust to the non-normality conditions. This finding is consistent with the previous studies examined non-normality of the level-1 and 2 (Owens & Farmer, 2013) and level-2 and 3 errors (Petit-Bois et al, 2013) in single-case data. Also, similar result was found in the general multilevel modeling literature examined non-normality of the level-2 errors using the robust standard error method (Mass & Hox, 2004).

However, the between-series model assuming homogeneous variances showed notably lower CI coverage rate than the nominal level (.95). Across simulation conditions, none of CI coverage rates were in the acceptable ranges and this pattern was consistent even for the condition where the level-1 errors were normally distributed. These results stand in contrast to previous research that showed nominal level coverage for the between-series model estimates when data were normally distributed (Ferron et al., 2014). Note that the between-series model initially examined in the current study used the homogeneous variance structure across the baseline and treatment phases, but the data generation model included not only baseline observation variance across participants (τ_{00}^2) but also treatment effect variance across participants (τ_{11}^2). Given that the between-series model has the flexibility to estimate a

heterogeneous variance structure for the baseline and treatment phases (Ferron et al., 2014), the follow-up study was conducted to evaluate the between-series model under the non-normality conditions when the heterogeneous variance model is specified.

The results of the follow-up study showed that the CI coverage rate was in the acceptable range across simulation conditions, and consistent with the previous results, skewness and kurtosis of the level-1 errors still did not affect CI coverage rate. CI coverage rate from the follow-up study was comparable to that of the within-series model. This result implies that reliable CI coverage rate can still be obtained using the between-series model for the non-normal level-1 error conditions if the model is correctly specified. For ReML and Bayesian inference methods, no significant difference was found in CI coverage rate. This result implies that confidence intervals using the Kenward-Roger method are comparable to HPD intervals.

Consistent with CI coverage rate results, CI widths results also showed no impact of the skewness and kurtosis the level-1 errors. However, wider CI width was consistently found from the between-series model compared to the within-series model. Furthermore, as the number of measurement occasions and participants increased, CI width became narrower. This finding implies that standard error decreased as the number of measurement occasions and participants increased. The same pattern regarding CI width can also be found in the previous studies with single-case simulation studies (e.g., Moeyaert et al., 2013a, 2014).

Statistical Power and Type I Error Rate of the Treatment Effect Estimate

In terms of statistical power to detect the treatment effect estimate, the within- and between-series models were robust to the non-normality of the level-1 error variance. Similar to the previous results, no distinct pattern was found across skewness and kurtosis conditions.

Statistical power for the within-series model showed considerably higher than the between-series model. In addition, statistical power for the within-series model increased as both number of measurement occasions and participants increased, whereas that of the between-series model increased only when the number of participants increased. Note these findings are consistent with the previous study (Ferron et al., 2014). Similar to CI coverage rate and width, ReML and Bayesian methods did not show a significant difference in power across simulation conditions.

With regards to Type I error, the within-series model showed well-controlled Type I error across skewness and kurtosis conditions. Type I error rates for the within-series model remained close to the nominal level (.05) as both skewness and kurtosis increased. This result is comparable to the results from the correctly specified within-series model as shown in previous studies (e.g., Ferron et al., 2010). However, similar to the CI coverage rate and width results, the between-series model with homogeneous variance showed the inflated Type I error rate. Type I error rate was around .10 across simulation conditions. The follow-up study results, however, showed that the between-series model with heterogeneous variance structure well-controlled Type I error rate across non-normality conditions. The result is comparable to the previous study with the correctly specified between-series model (Ferron et al., 2014).

ReML and Bayesian approaches also showed the similar result for Type I error control. Both approaches showed considerably well-controlled Type I error rates ranged from .05 to .07 across conditions. Note that previous studies compared ReML and Bayesian estimations for single-case context (e.g., Moeyaert et al., 2016) or multilevel modeling in general (e.g., Baldwin & Fellingham, 2013) did not compare the Type I error or statistical power results.

Bias and RMSE of the Parameters other than the Treatment Effect Estimate

In the current study, the bias and RMSE for the parameters other than the treatment effect were computed. For the within-series model, bias and RMSE of the fixed effect parameters (i.e., intercept [γ_{00}], time trend in baseline phase [γ_{20}], and change in time trend with treatment [γ_{30}]), level-1 error variance (σ^2), and level-2 error variance for the intercept (τ_{00}^2) and treatment effect (τ_{11}^2) were computed. For the between-series model, bias and RMSE of the level-1 error variance (σ^2) were computed.

Overall, skewness and kurtosis of the level-1 error variance did not affect the accuracy of the other fixed effect estimates in the within-series model. Relative bias results also showed less than 5% bias of the population parameter across conditions. Both ReML and Bayesian methods estimated the other fixed effect parameters with minimal bias across simulation conditions. As expected, RMSE of the fixed effect estimates decreased as both number of measurement occasions and participants increased. The bias and RMSE values for the fixed effect parameters were similar to the previous within-series single-case studies (e.g., Ferron et al., 2009, 2010; Moeyaert et al., 2013a, 2013b).

For the variance component estimates, as previous studies have shown, the estimates had some bias, especially for the level-2 error variances (Moyeart et al., 2013a, 2013b). The within-series model, however, showed relatively smaller bias for the level-1 error variance than the between-series model that assumed homogeneous variance. This pattern was consistent for the normal and non-normal level-1 error variance conditions. For the level-2 error variance estimates of the within-series model, substantial bias and relative bias were found. However, interestingly, the Bayesian method produced consistently smaller RMSE for the level-2 error variances for the

intercept and treatment effect parameters in the within-series model. This finding can also be found in the previous study with Bayesian single-case research (Moeyaert et al., 2016).

In sum, the current study found that estimation and inference for the treatment effect using the within- and between-series models are robust to non-normality of the level-1 errors. Minimal bias and relative bias of the treatment effect estimates were observed for both models. CI coverage rate was in the acceptable ranges with the within-series model and the heterogeneous variance between-series model. Also, Type I error was well-controlled as the normality of level-1 error was violated. The parameters other than the treatment effect in the within-series model were also accurately estimated across simulation conditions. Finally, no substantial difference was found between ReML with Kenward-Roger and Bayesian approaches for the fixed effect estimates, whereas the Bayesian method outperformed ReML for the level-2 error variance estimation.

Implications

In practice, it is not uncommon to observe non-normality of data from MB studies. Applied single-case researchers (e.g., Parker, 1996; Solomon, 2014) and methodologists (e.g., Shadish, 2014) have commented on this potentially problematic aspect of single-case data. However, there has been only limited research to systematically verify the robustness of statistical modeling approaches when data are not normally distributed. Owens and Farmer (2013) examined the robustness to non-normality of the level-1 and 2 errors for the within-series model using ReML. However, the robustness to non-normality for Bayesian estimation of single-case research or the robustness of the between series model has not been investigated previously. For this reason, this study was designed to provide practical information about how skewness

and kurtosis of level-1 errors impact the accuracy of estimation and statistical inference for the treatment effect parameter in within- and between-series models. In addition, this study investigated how various conditions including the numbers of participants and measurement occasions, and estimation methods (e.g., ReML and Bayesian) impact the performance of the models. The results lead to several implications for applied single-case researchers who are interested in the effect of interventions in MB studies, as well as for the methodologists who are interested in statistical methods for analyzing MB studies.

Implications for the Applied Researchers

The current study has following implications for the applied single-case researchers. First, it was found that both within- and between-series models are robust to the non-normality in level-1 errors. One of the concerns that may arise with empirical MB data is the violation of the normality assumption due to the scale of the measurement or outliers. However, based on the results of the study, the within- and between-series models can estimate the parameters of the models with precision even though the normality assumption is violated. Minimal bias of the treatment effect estimate was observed, for the situations where skewness and kurtosis of the level-1 errors ranged from 0 to 3 and -1 to 4, respectively. In addition, the other fixed effects in the within-series model including intercept, time trend effects in baseline and change in time trend with treatment were also estimated without substantial bias. These findings imply that when applied researchers want to estimate treatment effects in their MB studies that the within- and between-series models can be used even when the outcomes are non-normally distributed with skewness in the level-1 errors as high as 3 and kurtosis as high as 4.

Second, statistical inferences about the treatment effect in MB studies were often reliable using the within- and between-series models, although the normality assumption is violated. More specifically, acceptable CI coverage rates and well-controlled Type I error rates were observed for the within-series model and the between-series model when the models were correctly specified. Similarly, statistical power was not affected by the non-normality of level-1 errors. However, the results of the study showed that unacceptable CI coverage rates and inflated Type I error rates can be found if the variance structure of the between-series model is misspecified. Although no substantial bias was observed for the between-series model with homogeneous and heterogeneous variance structures, it is recommended that care be taken to specify the variance structure accurately to obtain reliable statistical inference. More specifically, if an applied researcher expects the treatment effect to vary across participants, the heterogeneous variance structure across the baseline and treatment phases for the between-series model is recommended.

Third, parameter estimation and inference of the within- and between-series models were accurate with relatively small numbers of participants and measurement occasions. Based on the study results, minimal bias and acceptable inference of the parameter estimates were still observed with as low as four participants, and ten measurement occasions from both within- and between-series models. In MB studies, it is common to find a relatively small number of participants, and as a result, the accuracy of parameter estimates may concern researchers. Based on the study results, it is encouraging that minimal bias and reliable inference can still be obtained using the within- and between-series model under small sample size conditions.

Implications for the Methodologists

The current study compared the performances of the widely-used within-series model and the newly-proposed between-series model under various non-normality conditions. Based on the results, correctly specified within- and between-series models are comparable in terms of parameter estimation bias and accuracy of statistical inferences. The between-series model, however, suffers from relatively lower statistical power and higher RMSE of parameter estimates due to the small sample size. However, as Ferron et al. (2014) shown, the between-series model effect estimates were not biased in circumstances where the within-series model effect estimates were, particularly when the within-series model is misspecified as a result of event effects, such as history, maturation, or instrumentation. By comparing the average treatment effect estimates from the between- and within-series models researchers can potentially detect model misspecification (Ferron et al., 2014). In addition, methodologists should also consider averaging the treatment effect at multiple time points (e.g., 1, 2, and 3 observations after treatment) to get more stability in the between-series estimate. Although in the current study, only one time point after treatment was considered for the between-series model, one can obtain a more stable estimate if multiple time point estimates were averaged. Ferron et al. (2014) showed that the multiple time point between-series estimate has lower RMSE and higher power than the single time point estimate. However, methodologists should also recognize issues with the multiple time point approach including additional complexity of the model and practical limitations of handling delayed treatment effects or limited time between intervention start points.

The current study also compared two estimation methods, ReML, and Bayesian estimations. ReML is often used in practice because a number of statistical programs used ReML as a default and it produces the results considerably faster than Bayesian estimation. Bayesian

estimation, however, also has an advantage for estimating more complex models because it incorporates prior distributions for the parameters of the model and sampling methods to obtain posterior distributions. Interestingly, both ReML and Bayesian estimations showed indistinguishable results for the fixed effect estimates. Both methods estimated the fixed effects accurately and their inferences were identically precise. This finding implies that both ReML and Bayesian methods are robust to the non-normality in the level-1 errors and they estimate fixed effect parameters comparably well using the within- and between-series models. Given that the current study used non-informative priors for the parameters in the models, this finding also implies that non-informative prior of Bayesian estimation yields the identical fixed effect estimates as ReML estimation. Although non-informative prior and ReML fixed effect estimates were robust to the non-normality in level-1 error, methodologists should also consider informative priors for the treatment effect in MB studies to increase precision. It is possible that one can collect information about means and standard deviations or lower and upper boundaries of the previous MB treatment effect estimates to determine the most appropriate informative priors.

From the results of the study, the variance component estimates showed some differences. As previous studies have shown, the variance components of the within-series model have are biased using ReML estimation (e.g., Ferron et al., 2009, 2010; Moeyaert et al., 2013a, 2013b). Bayesian estimation has been considered as an alternative approach for the variance component estimation in the context of single-case research because it allows researchers to specify a prior distribution for the variance component (Moeyaert et al., 2016). From the current study results, it was found that Bayesian estimation yielded more accurate estimates for the level-2 variance components. Lower RMSE of the level-2 error variance for intercept and

treatment effect in the within-series model was observed. Because of the prior specification in the level-2 error distribution, Bayesian estimation results in better parameter estimates than ReML when the sample size is limited. This finding suggests that Bayesian estimation should be encouraged if a researcher is interested in estimating the level-2 error variance in single-case research data. For example, if a researcher is interested in obtaining the standardized effect size using the within- or between-series model, then less biased variance estimates should be used and MCMC might be the more appropriate approach. In addition, methodologists should also consider the small sample bias correction if the standardized effect size is of primary interest because biased variance components tend to result in biased effect size estimates (Hedges, Pustejovsky, & Shadish, 2012). Hedges et al. (2012) provides the small sample correction approach for the standardized effect size in single-case research design.

Limitation and Future Research

There are several limitations in the current study. First, the study was conducted with only limited ranges of the skewness and kurtosis of the level-1 error variance. As the results of the study indicated, the within- and between-series estimators are robust to the degrees of skewness and kurtosis examined. Those degrees of the skewness and kurtosis were chosen based on a preliminary survey of published single-case research design data and also can be considered as probable ranges for applied settings. However, it is possible for single-case research design data to have more extreme degrees of skewness and kurtosis of the level-1 errors. The results may be different if more extreme skewness or kurtosis of the level-1 errors was present. It is worthwhile to investigate the performance of the within- and between-series models with conditions where the level-1 errors were generated with more extreme skewness and kurtosis.

Second, non-normality conditions in the current study design were only manipulated with degrees of skewness and kurtosis assuming dependent variables are continuous. Non-normality, however, may also occur if the scales of measurement are not continuous variables including counts, proportions or percentages (Shadish et al., 2013). Past research has suggested using more complex and sophisticated statistical models to fit those types of non-normal data. However, limited studies investigated the efficacy of more complex statistical models when non-continuous MB data are distributed as non-normal. Since fitting more complex models to MB studies could create potential problems such as estimation with small sample sizes (Shadish, Kyse, & Rindskopf, 2013) or misspecification of the underlying population distribution (Shadish, Zuur, & Sullivan, 2014), more empirical-based research is needed to examine the efficiency of more complex models to non-normal MB data.

Third, the current study investigated non-normality in only level-1 errors. Non-normality in level-2 errors may occur in practical situations and only limited studies investigated this issue (e.g., Owens & Farmer, 2013; Petit-Bois et al., 2013). The comparison of the within- and between-series models in conjunction with non-normality in the level-2 error variances has not been investigated. In addition, given that Bayesian estimation showed better results in estimating the level-2 error variance in the current study, it would be interesting to examine ReML and Bayesian estimations for the level-2 error variance when non-normality occurs in level-2 errors. Future research can examine the performances of various statistical approaches by creating non-normality in level-2 error distributions.

Lastly, when the within- and between-series models were estimated with Bayesian estimation, only non-informative prior distributions were considered. As the current study showed, similar results in the fixed effect estimates from ReML and Bayesian estimation were

found using non-informative prior distributions. A similar pattern was observed from the previous studies as well (e.g., Baek, 2015; Moeyaert et al., 2016). Informative prior distributions can be obtained from the meta-analysis of MB studies or survey of the published MB data. It is also possible to assume different underlying prior distributions such as a student's t or Cauchy distribution for the fixed effect parameter, and uniform distribution, or (not inverse) Wishart distribution for the variance component. More research is needed to examine the performance of different prior distributions in Bayesian modeling with MB studies under various situations.

Nonetheless, the results of this study provide valuable information about how to deal with non-normal MB data using the within- and between-series models, and ReML with Kenward-Roger and Bayesian methods, which are being considered for single-case research. The author hopes the study will allay concerns about the usefulness of the within- and between-series treatment effect estimators and in doing so encourage the estimation and reporting of effect estimates MB research.

REFERENCES

Baek, E. K. (2015). *Consequences of Non-Modeled and Modeled Between Case Variation in the Level-1 Error Structure in Multilevel Models for Single-Case Data: A Monte Carlo Study*. (Unpublished doctoral dissertation). University of South Florida: Tampa, FL.

Baek, E.K., & Ferron, J. M. (2013). Multilevel models for multiple-baseline data: Modeling across participant variation in autocorrelation and residual variance. *Behavior Research Methods*, 45, 65-74.

Baldwin, S. A., & Fellingham, G. W. (2013). Bayesian methods for the analysis of small sample multilevel data with a complex variance structure. *Psychological Methods*, *18*, 151-164.

Barlow, D. H., Nock, M. K., & Hersen, M. (2009). *Single Case Experimental Designs: Strategies for Studying Behavior Change* (3rd ed.). Boston, MA: Pearson.

Bell, B. A., Morgan, G. B., Schoeneberger, J. A., Kromrey, J. D., & Ferron, J. M. (2014). How low can you go? *Methodology*, *10*, 1-11.

Beretvas, S. N., & Chung, H. (2008). A review of meta-analyses of single-subject experimental designs: Methodological issues and practice. *Evidence-Based Communication Assessment and Intervention*, *2*, 129-141.

Browne, W. J., & Draper, D. (2000). Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. *Computational Statistics*, *15*, 391-420.

Browne, W. J., & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, *1*, 473-514.

Browne, W. J., Draper, D., Goldstein, H., & Rasbash, J. (2002). Bayesian and likelihood methods for fitting multilevel models with complex level-1 variation. *Computational Statistics & Data Analysis*, *39*, 203-225.

Bunuan, R., Hembry, I. & Beretvas, S.N. (2013). *Meta-analysis of single-case design studies with multiple baselines across participants and settings.* Unpublished Manuscript.

Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine*, *25*, 4279-4292.

Campbell, D. T., & Stanley, J. C. (2063). *Experimental and quasi-experimental designs for research*. Ravenio Books.

Clarke, P., & Wheaton, B. (2007). Addressing data sparseness in contextual population research using cluster analysis to create synthetic neighborhoods. *Sociological Methods & Research, 35,* 311-351.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.

Crawford, J. R., Garthwaite, P. H., Azzalini, A., Howell, D. C., & Laws, K. R. (2006). Testing for a deficit in single-case studies: Effects of departures from normality. *Neuropsychologia, 44*, 666-677.

de Vries, R. M., & Morey, R. D. (2013). Bayesian hypothesis testing for single-subject designs. *Psychological Methods, 18,* 165-185.

DiCarlo, C. F., & Reid, D. H. (2004). Increasing pretend toy play of toddlers with disabilities in an inclusive setting. *Journal of Applied Behavior Analysis, 37,* 197-207.

Edgington, E., & Onghena, P. (2007). *Randomization tests*. CRC Press.

Edwards, T. L., La Londe, K. B., Cox, C., Weetjens, B., & Poling, A. (2015). Effects of schedules of reinforcement on pouched rats' performance in urban search-and-rescue training. *Journal of Applied Behavior Analysis*, 49, 199-204.

Eliason, S. R. (1993). *Maximum likelihood estimation: Logic and practice* (Vol. 96). Sage Publications.

Farmer, J., Owens, C. M., Ferron, J. M. and Allsopp, D. (2010, May). *A review of social science single-case meta-analyses.* Paper presented at the annual meeting of the American Education Research Association, Denver, CO.

Ferron, J. M., Bell, B. A., Hess, M. R., Rendina-Giobioff, G. & Hibbard, S.T. (2009). Making treatment effect inferences from multiple baseline data: The utility of multilevel modeling approaches. *Behavior Research Methods, 41,* 372-384.

Ferron, J. M., Farmer, J. L., & Owens, C. M. (2010). Estimating individual treatment effects from multiple-baseline data: A Monte Carlo study of multilevel-modeling approaches. *Behavior Research Methods, 42,* 930-943.

Ferron, J. M., Foster-Johnson, L., & Kromrey, J. D. (2003). The functioning of single-case randomization tests with and without random assignment. *The Journal of Experimental Education, 71,* 267-288.

Ferron, J. M., & Jones, P. K. (2006). Tests for the visual analysis of response-guided multiple-baseline data. *The Journal of Experimental Education, 75*, 66-81.

Ferron, J. M., Joo, S. H., & Levin, J. R. (2016). *A Monte-Carlo evaluation of masked-visual analysis in response-guided versus fixed-criteria multiple-baseline designs.* Unpublished Manuscript.

Ferron, J. M., & Levin, J. R. (2014). Single-case permutation and randomization statistical tests: Present status, promising new developments. In T. R. Kratochwill & J. R. Levin (Eds.) *Single-case intervention research: Methodological and statistical advances.* Washington, D.C.: American Psychological Association.

Ferron, J. M., Moeyaert, M., Van den Noortgate, W., & Beretvas, S. N. (2014). Estimating causal effects from multiple-baseline studies: Implications for design and analysis. *Psychological Methods, 19*, 493-510.

Ferron, J., & Onghena, P. (1996). The power of randomization tests for single-case phase designs. *The Journal of Experimental Education, 64*, 231-239.

Ferron, J., & Sentovich, C. (2002). Statistical power of randomization tests used with multiple-baseline designs. *The Journal of Experimental Education, 70*, 165-178.

Ferron, J. M., & Scott, H. (2005). Multiple baseline designs. *Wiley Stats Ref: Statistics Reference Online*.

Ferron, J., & Ware, W. (1995). Analyzing single-case data: The power of randomization tests. *The Journal of Experimental Education, 63*, 167-178.

Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika, 43*, 521-532.

Fisch, G. S. (2001). Evaluating data from behavioral analysis: Visual inspection or statistical models?. *Behavioural Processes, 54*, 137-154.

Fisher, W. W., Kelley, M. E., & Lomas, J. E. (2003). Visual aids and structured criteria for improving visual inspection and interpretation of single-case designs. *Journal of Applied Behavior Analysis, 36*, 387-406.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis, 1*, 515-534.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 2). London: Chapman & Hall/CRC.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 457-472.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In Bayesian Statistics 4 (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) 169-193. Oxford Univ. Press.

Goldstein, H. (1995). Hierarchical data modeling in the social sciences. *Journal of Educational and Behavioral Statistics*, *20*, 201-204.

Gottman, J. M., & Glass, G. V. (1978). Analysis of interrupted time-series experiments. In T. R. Kratochwill (Ed.), *Single subject research: Strategies for evaluating change* (pp. 197-234). Academic Press New York.

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods*, *3*, 224-239.

Hembry, I., Bunuan, R., Beretvas, S. N., Ferron, J. M., & Van den Noortgate, W. (2015). Estimation of a nonlinear intervention phase trajectory for multiple-baseline design data. *The Journal of Experimental Education*, *83*, 514-546.

Heidelberger, P., & Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, *31*, 1109-1144.

Himle, M. B., & Wright, K. A. (2014). Behavioral skills training to improve installation and use of child passenger safety restraints. *Journal of Applied Behavior Analysis*, *47*, 549-559.

Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling An overview and a meta-analysis. *Sociological Methods & Research*, *26*, 329-367.

Hornor, R. H., & Odom, S. L. (2014). Constructing Single-Case Research Designs: Logic and Options. In T. R. Kratochwill & J. R. Levin (Eds.) *Single-Case Intervention Research: Methodological and Statistical Advances.* Washington D.C.: American Psychological Association.

Hox, J. (1998). Multilevel modeling: When and why. In *Classification, data analysis, and data highways* (pp. 147-154). Springer Berlin Heidelberg.

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2010). *Multilevel analysis: Techniques and applications*. Routledge.

Huitema, B. E., & Mckean, J. W. (2000). Design specification issues in time-series intervention models. *Educational and Psychological Measurement*, *60*, 38-58.

Hunt, P., Soto, G., Maier, J., & Doering, K. (2003). Collaborative teaming to support students at risk and students with severe disabilities in general education classrooms. *Exceptional Children*, *69*, 315-332.

Ittenbach, R. F., & Lawhead, W. F. (1997). Historical and philosophical foundations of single-case research. In R. D. Franklin, D. B. Allison, & B. S. Gorman (Eds.), *Design and analysis of single-case research* (pp. 13–39). Mahwah, NJ: Erlbaum.

Joo, S. H., Ferron, J. M., Moeyaert, M., Beretvas, S. N., & Van den Noortgate, W. (2016). *Approaches for specifying the level-1 error structure when synthesizing single-subject data,* Unpublished Manuscript.

Kazdin, A. E. (2011). *Single-Case Research Designs: Methods for Clinical and Applied Settings* (2nd ed.). New York: Oxford University Press.

Kazdin, A. E., & Kopel, S. A. (1975). On resolving ambiguities of the multiple-baseline design: Problems and recommendations. *Behavior Therapy, 6*, 601-608.

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53*, 983-997.

Koehler, M. J., & Levin, J. R. (1998). Regulated randomization: A potentially sharper analytical tool for the multiple-baseline design. *Psychological Methods, 3*, 206-217.

Koutsoftas, A. D., Harmon, M. T., & Gray, S. (2009). The effect of tier 2 intervention for phonemic awareness in a response-to-intervention model in low-income preschool classrooms. *Language, Speech, and Hearing Services in Schools, 40*, 116-130.

Kratochwill, T. R. (Ed.). (1978*). Single subject research: Strategies for evaluating change.* New York, NY: Academic Press.

Kratochwill, T. R. (1985). Case study research in school psychology. *School Psychology Review, 14*, 204–215.

Kratochwill. T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M. & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education, 34, 26-38.*

Kratochwill, T. R., & Levin, J. R. (Eds.). (1992). *Single-case research design and analysis: New directions for psychology and education.* Hillsdale, NJ: Erlbaum.

Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods, 15*, 124-144.

Kratochwill, T. R., Levin, J. R., Horner, R. H., & Swoboda, C. M. (2014). Visual analysis of single-case intervention research: Conceptual and methodological issues. In T. R. Kratochwill & J. R. Levin (Eds.) *Single-Case Intervention Research: Methodological and Statistical Advances.* Washington D.C.: American Psychological Association.

Langford, I. H., & Lewis, T. (1998). Outliers in multilevel data. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 161*, 121-160.

Ma, H. H. (2006). An alternative method for quantitative synthesis of single-subject researches percentage of data points exceeding the median. *Behavior Modification, 30,* 598-617.

Maas, C. J., & Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica, 58,* 127-137.

Maggin, D. M., Swaminathan, H., Rogers, H. J., O'Keeffe, B. V., Sugai, G., & Horner, R. H. (2011). A generalized least squares regression approach for computing effect sizes in single-case research: Application examples. *Journal of School Psychology, 49,* 301-321.

Marascuilo, L. A., & Busk, P. L. (1988). Combining statistics for multiple-baseline AB and replicated ABAB designs across subjects. *Behavioral Assessment, 10,* 1-28

Mastropieri, M. A., & Scruggs, T. E. (1985). Early intervention for socially withdrawn children. *The Journal of Special Education, 19,* 429-441.

Matyas, T. A., & Greenwood, K. M. (1990). Visual analysis of single-case time series: Effects of variability, serial dependence, and magnitude of intervention effects. *Journal of Applied Behavior Analysis, 23,* 341-351.

Moeyaert, M., Rindskopf, D., Onghena, P., & Van den Noortgate, W. (2016). *Multilevel modeling of single-case data: A comparison of Maximum Likelihood and Bayesian Estimation*, Unpublished Manuscript.

Moeyaert, M., Ugille, M., Ferron, J. M., Beretvas, S. N., & Van den Noortgate, W. (2013a). The three-level synthesis of standardized single-subject experimental data: A Monte Carlo simulation study. *Multivariate Behavioral Research, 48,* 719-748.

Moeyaert, M., Ugille, M., Ferron, J. M., Beretvas, S. N., & Van den Noortgate, W. (2013b). Modeling external events in the three-level analysis of multiple-baseline across-participants designs: A simulation study. *Behavior Research Methods, 45,* 547-559.

Moeyaert, M., Ugille, M., Ferron, J. M., Beretvas, S. N., & Van den Noortgate, W. V. (2014). Three-level analysis of single-case experimental data: Empirical validation. *The Journal of Experimental Education, 82,* 1-21.

Moeyaert, M., Ugille, M., Ferron, J. M., Beretvas, S. N., & Van den Noortgate, W. (2016). The Misspecification of the Covariance Structures in Multilevel Models for Single-Case Data: A Monte Carlo Simulation Study. *The Journal of Experimental Education, 84*, 473-509.

Onghena, P. (2005). Single-case designs. In B. Everitt & D. Howell (Eds.), *Encyclopedia of statistics in behavioral science: Vol. 4* (pp. 1850-1854). New York, NY: Wiley.

Owens, C. M., & Farmer, J. L. (2013, April). *Analyzing multiple baseline data using multilevel modeling with various residual distributions: A Monte Carlo study.* Paper presented at the American Educational Research Association Annual Meeting, San Francisco, CA.

Parker, R. I., (2006). Increased reliability for single case research results: Is the bootstrap the answer? *Behavior Therapy*, *37*, 326-338.

Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percentage of all non-overlapping data (PAND) an alternative to PND. *The Journal of Special Education*, *40*, 194-204.

Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy*, *40*, 357-367.

Parker, R. I., Vannest, K. J., & Brown, L. (2009). The improvement rate difference for single-case research. *Exceptional Children*, *75*, 135-150.

Parker, R. I, Vannest, K. J., & Davis, J. L. (2014). Non-overlap analysis for single-case research. In T. R. Kratochwill & J. R. Levin (Eds.) *Single-Case Intervention Research: Methodological and Statistical Advances.* Washington D.C.: American Psychological Association.

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy*, *42*, 284-299.

Petit-Bois, M. (2014). *A Monte Carlo study: The consequences of the misspecification of the level-1 error structure*. (Unpublished doctoral dissertation). University of South Florida: Tampa, FL.

Petit-Bois, M., Baek, E. K., Nguyen, D., & Ferron, J. M. (2013, April). *The effect of error structure specification on the meta-analysis of single-case studies: A Monte Carlo study*. Poster presented at the American Educational Research Association Annual Meeting, San Francisco, CA.

Petit-Bois, M., Baek, E. K., Van den Noortgate, W., Beretvas, S. N., & Ferron, J. M. (2016). The consequences of modeling autocorrelation when synthesizing single-case studies using a three-level model. *Behavior Research Methods*, *48*, 803-812.

Raftery, A. E., & Lewis, S. (1992). How many iterations in the Gibbs sampler. *Bayesian Statistics*, *4*, 763-773.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and Data Analysis Methods* (Vol. 1). Sage.

Rindskopf, D. (2014). Nonlinear Bayesian analysis for single case designs. *Journal of School Psychology*, *52*, 179-189.

Rindskopf, D. M., & Ferron, J. M. (2014). Using multilevel models to analyze single-case design data. In T. R. Kratochwill & J. R. Levin (Eds.) *Single-Case Intervention Research: Methodological and Statistical Advances.* Washington D.C.: American Psychological Association.

SAS Institute. (2014). *SAS 9.4,* Computer Software. SAS Inst.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin,* 110-114.

Scotti, J. R., Evans, I. M., Meyer, L. H., & Walker, P. (1991). A meta-analysis of intervention research with problem behavior: Treatment validity and standards of practice. *American Journal on Mental Retardation*, *96,* 233-256.

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research methodology and validation. *Remedial and Special Education, 8,* 24-33.

Shadish, W. R. (2014). Analysis and meta-analysis of single-case designs: An introduction. *Journal of School Psychology, 52,* 109-122.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Houghton, Mifflin and Company.

Shadish, W.R., Kyse, E.N., & Rindskopf, D.M. (2013). Analyzing data from single-case designs using multilevel models: New applications and some agenda items for future research. *Psychological Methods*, *18,* 385-405.

Shadish, W. R., & Rindskopf, D. M. (2007). Methods for evidence-based practice: Quantitative synthesis of single-subject designs. *New Directions for Evaluation*, *2007,* 95-109.

Shadish, W. R., Rindskopf, D. M., Hedges, L. V., & Sullivan, K. J. (2013). Bayesian estimates of autocorrelations in single-case designs. *Behavior Research Methods*, *45,* 813-821.

Shadish, W. R., & Sullivan, K. J. (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research Methods*, *43,* 971-980.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52,* 591-611.

Smith, J. D. (2012). Single-case experimental designs: A systematic review of published research and current standards. *Psychological Methods*, *17,* 510-550.

Solomon, B. G. (2014). Violations of assumptions in school-based single-case data: Implications for the selection and interpretation of effect sizes. *Behavior Modification, 38,* 477-496.

Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). *WinBUGS User Manual 1.4*. Retrieved from http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf.

Swaminathan, H., Rogers, H. J., & Horner, R. H. (2014). An effect size measure and Bayesian analysis of single-case designs. *Journal of School Psychology, 52,* 213-230.

Van den Noortgate, W., & Onghena, P. (2003a). Combining single-case experimental data using hierarchical linear models. *School Psychology Quarterly, 18*, 325-346.

Van den Noortgate, W., & Onghena, P. (2003b). Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavior Research Methods, Instruments, & Computers*, *35*, 1-10.

Van den Noortgate, W., Onghena, P. (2007). The aggregation of single-case results using hierarchical linear models. *The Behavior Analyst Today, 8*, 196-209.

Van den Noortgate, W., & Onghena, P. (2008). A multilevel meta-analysis of single-subject experimental designs. *Evidence-Based Communication Assessment and Intervention, 2*, 142-151.

Van der Leeden, R., Busing, F. M. T. A., & Meijer, E. (1997, April). *Applications of bootstrap methods for two-level models,* Paper presented at the Multilevel Conference, Amsterdam, Netherland.

Vargo, K. K., & Ringdahl, J. E. (2015). An evaluation of resistance to change with unconditioned and conditioned reinforcers. *Journal of Applied Behavior Analysis, 48*, 643-662.

Wacker, D. P., Steege, M., & Berg, W. K. (1988). Use of single-case designs to evaluate manipulable influences on school performance. *School Psychology Review*, *17*, 651–657.

Wampold, B. E., & Worsham, N. L. (1986). Randomization tests for multiple-baseline designs. *Behavioral Assessment, 8,* 135-143.

Washington, W. D., Banna, K. M., & Gibson, A. L. (2014). Preliminary efficacy of prize-based contingency management to increase activity levels in healthy adults. *Journal of Applied Behavior Analysis, 47,* 231-245.

APPENDIX A. TABLES OF THE COMPLETE RESULTS

Table 5.
Relative Bias for the Treatment Effect

		I=10								*I=20*								*I=40*							
		J = 4				*J = 8*				*J = 4*				*J = 8*				*J = 4*				*J = 8*			
		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC	
Skew	Kurt	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B
0	-1	.026	.023	.024	.023	-.015	-.014	-.018	-.014	-.012	-.005	-.013	-.006	.009	-.004	.014	-.004	-.003	.008	-.004	.008	-.011	-.028	-.007	-.028
	0	-.021	-.040	-.016	-.040	-.003	.002	-.005	.002	.009	.016	.014	.015	.001	.010	-.003	.010	.001	.012	-.007	.013	.006	.017	.003	.017
	1	.004	-.003	.016	-.003	.003	-.001	.005	.000	.001	.014	-.002	.014	.005	-.001	.002	-.001	-.005	-.024	-.008	-.024	.005	-.009	.005	-.009
	2	-.005	.011	-.011	.011	.001	-.008	.001	-.007	.003	.018	.004	.018	-.008	-.008	-.010	-.008	.005	-.011	.002	-.011	-.006	-.008	-.001	-.008
	4	.010	.000	-.008	.000	-.008	-.003	-.008	-.004	.004	-.010	.004	-.010	-.001	.020	.001	.020	.005	.009	.003	.009	.003	-.002	.003	-.002
1	-1	-.001	-.004	.001	-.005	.000	-.004	-.006	-.004	-.003	.013	-.001	.013	-.014	-.012	-.016	-.012	.002	-.003	.002	-.002	-.002	.004	.001	.004
	0	.013	.000	.009	.000	.000	.002	.000	.002	-.009	.006	-.014	.006	.001	-.010	.001	-.010	-.016	.025	-.024	.024	.010	-.005	.010	-.005
	1	.021	.041	.017	.041	.017	.019	.008	.019	.005	.004	.005	.004	.001	-.014	.002	-.014	.010	-.011	.005	-.011	.012	.020	.009	.020
	2	-.023	-.031	-.008	-.030	.004	.018	-.005	.018	-.015	-.031	-.016	-.031	-.010	.000	-.010	.000	.008	.011	.004	.010	-.016	-.002	-.019	-.003
	4	-.001	.011	-.002	.012	.015	.013	.014	.012	.016	.029	.008	.029	-.002	.015	.001	.014	-.008	-.010	-.007	-.010	-.009	-.010	-.010	-.010
2	-1	-.005	.002	-.007	.002	-.004	.005	-.001	.005	-.022	-.019	-.018	-.020	-.011	-.023	-.010	-.023	.000	.003	-.003	.003	.003	.007	.001	.008
	0	.003	-.012	-.002	-.012	.014	.013	.009	.012	-.001	.008	-.005	.008	-.002	.002	-.007	.002	-.007	-.010	-.011	-.010	.000	-.006	-.005	-.006
	1	-.041	-.033	-.025	-.032	-.017	.006	-.013	.006	-.004	.016	.001	.015	-.005	.001	-.003	.001	-.012	.013	-.021	.013	.005	.013	.005	.013
	2	.010	-.006	.002	-.005	-.005	-.007	-.002	-.007	.018	.027	.008	.027	-.003	.008	-.001	.008	-.007	-.020	-.016	-.021	-.004	-.014	-.003	-.014
	4	.030	-.011	.030	-.012	.006	.010	-.001	.010	-.012	-.023	-.016	-.022	-.007	-.015	-.007	-.015	-.006	-.015	-.008	-.015	.001	-.005	.002	-.005
3	-1	.007	.010	-.007	.010	.018	.007	.023	.006	.000	.021	-.004	.021	-.006	-.023	-.002	-.023	.005	.015	.005	.016	.003	-.002	.002	-.002
	0	.009	.022	.016	.022	-.014	-.019	-.009	-.019	-.018	-.009	-.013	-.008	.015	-.011	.018	-.012	.002	.020	-.004	.020	.006	.003	.002	.003
	1	.015	-.009	.008	-.009	.009	.013	.002	.014	.012	.025	.008	.025	-.001	.003	-.006	.003	-.013	-.032	-.008	-.032	-.005	-.015	-.004	-.015
	2	-.002	.001	-.007	.001	.007	.014	.004	.014	-.012	-.021	-.009	-.021	.001	-.002	.006	-.002	-.002	.011	-.008	.012	.003	.004	.004	.004
	4	.003	-.015	.012	-.015	-.007	.009	-.017	.009	.001	.000	-.001	.000	.002	-.005	.000	-.005	.003	.010	-.001	.010	-.011	-.019	-.010	-.018

Note. Skew = skewness of the level-1 error variance, Kurt = kurtosis of the level-1 error variance, ReML = restricted maximum likelihood estimation, MCMC = Markov chain Monte Carlo estimation, W = within-series model, B = between-series model, *I* = number of measurement occasions, *J* = number of participants.

Table 6.
RMSE for the Treatment Effect

		I=10								*I=20*								*I=40*							
		J = 4				*J = 8*				*J = 4*				*J = 8*				*J = 4*				*J = 8*			
		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC	
Skew	Kurt	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B
0	-1	.872	1.101	.967	1.102	.571	.742	.640	.742	.567	.940	.628	.940	.404	.676	.438	.675	.464	.967	.503	.967	.339	.697	.367	.697
	0	.842	1.073	.929	1.073	.574	.746	.653	.746	.572	.946	.622	.947	.401	.694	.437	.694	.464	.984	.505	.984	.327	.686	.354	.686
	1	.840	1.045	.935	1.046	.577	.753	.650	.753	.574	.956	.632	.956	.405	.680	.447	.680	.469	.970	.507	.971	.324	.672	.356	.673
	2	.846	1.079	.943	1.079	.567	.758	.636	.758	.577	.946	.629	.946	.400	.674	.445	.675	.463	.954	.504	.954	.316	.683	.345	.683
	4	.840	1.072	.939	1.072	.574	.758	.640	.758	.581	.959	.631	.959	.402	.686	.442	.685	.466	.960	.503	.960	.326	.679	.354	.679
1	-1	.848	1.103	.931	1.103	.576	.750	.649	.750	.590	.964	.645	.965	.405	.680	.445	.680	.476	.996	.518	.996	.330	.704	.359	.704
	0	.826	1.046	.924	1.046	.577	.744	.662	.745	.576	.945	.632	.945	.406	.687	.446	.688	.465	.973	.511	.973	.326	.683	.356	.683
	1	.847	1.064	.930	1.064	.575	.744	.650	.744	.567	.961	.626	.961	.404	.687	.443	.688	.463	.963	.506	.963	.321	.689	.354	.690
	2	.849	1.061	.949	1.061	.570	.732	.640	.732	.573	.956	.627	.957	.400	.680	.445	.681	.465	.975	.504	.975	.329	.675	.356	.676
	4	.837	1.043	.933	1.043	.567	.746	.630	.746	.582	.949	.640	.949	.402	.681	.439	.681	.466	.973	.506	.973	.335	.688	.365	.688
2	-1	.830	1.061	.931	1.061	.572	.745	.657	.746	.581	.969	.632	.968	.406	.679	.447	.679	.468	.955	.509	.955	.330	.688	.362	.688
	0	.831	1.065	.926	1.065	.573	.754	.647	.754	.581	.938	.630	.938	.397	.673	.440	.673	.470	.989	.505	.990	.326	.691	.354	.691
	1	.834	1.071	.932	1.071	.571	.743	.644	.744	.588	.954	.646	.954	.400	.680	.438	.680	.463	.965	.505	.966	.322	.684	.351	.684
	2	.854	1.070	.951	1.069	.577	.756	.649	.756	.585	.963	.628	.963	.400	.683	.439	.683	.457	.987	.498	.988	.330	.686	.359	.686
	4	.845	1.057	.943	1.057	.572	.759	.643	.760	.576	.961	.632	.962	.405	.688	.440	.688	.468	.986	.506	.986	.330	.667	.361	.667
3	-1	.843	1.089	.931	1.090	.579	.745	.653	.745	.584	.975	.643	.975	.405	.665	.443	.665	.464	.987	.505	.988	.324	.668	.354	.668
	0	.838	1.053	.920	1.054	.577	.757	.641	.757	.574	.969	.637	.969	.411	.686	.453	.686	.464	.973	.502	.974	.333	.696	.363	.696
	1	.821	1.032	.913	1.033	.581	.762	.644	.762	.585	.963	.640	.964	.400	.675	.444	.675	.463	.977	.506	.978	.326	.691	.354	.691
	2	.838	1.047	.941	1.047	.574	.742	.651	.743	.567	.952	.616	.953	.406	.672	.451	.672	.469	.969	.511	.970	.335	.701	.366	.702
	4	.835	1.054	.934	1.054	.566	.739	.640	.740	.568	.962	.616	.962	.399	.678	.436	.678	.465	.962	.511	.962	.324	.683	.351	.683

Note. Skew = skewness of the level-1 error variance, Kurt = kurtosis of the level-1 error variance, ReML = restricted maximum likelihood estimation, MCMC = Markov chain Monte Carlo estimation, W = within-series model, B = between-series model, *I* = number of measurement occasions, *J* = number of participants.

Table 7.

CI coverage Rate for the Treatment Effect

		I=10								*I=20*								*I=40*							
		J = 4				*J = 8*				*J = 4*				*J = 8*				*J = 4*				*J = 8*			
		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC	
Skew	Kurt	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B
0	-1	**.937**	**.867**	**.935**	**.857**	.951	**.884**	.943	**.879**	.942	**.882**	**.939**	**.870**	.947	**.871**	**.941**	**.869**	**.934**	**.864**	**.937**	**.863**	**.932**	**.858**	**.916**	**.855**
	0	.953	**.869**	.945	**.859**	.949	**.882**	.943	**.877**	.948	**.871**	.943	**.864**	.947	**.857**	**.934**	**.852**	**.935**	**.864**	**.929**	**.861**	.945	**.860**	**.935**	**.855**
	1	.950	**.878**	.944	**.872**	.947	**.881**	.943	**.875**	.946	**.878**	**.941**	**.873**	.948	**.872**	**.933**	**.867**	**.936**	**.861**	**.933**	**.856**	.947	**.875**	**.928**	**.873**
	2	.945	**.875**	**.940**	**.862**	.953	**.870**	.946	**.866**	.942	**.874**	**.939**	**.863**	.943	**.871**	**.935**	**.868**	**.934**	**.871**	**.936**	**.864**	.950	**.860**	**.934**	**.858**
	4	.944	**.869**	**.941**	**.859**	.953	**.876**	.948	**.870**	.945	**.875**	**.938**	**.869**	.944	**.868**	**.938**	**.864**	**.934**	**.865**	**.941**	**.857**	.947	**.870**	**.927**	**.867**
1	-1	.945	**.861**	.947	**.847**	.948	**.874**	.946	**.868**	.943	**.871**	**.930**	**.865**	.945	**.867**	**.935**	**.867**	**.928**	**.869**	**.929**	**.862**	.943	**.856**	**.923**	**.853**
	0	.952	**.878**	.942	**.868**	.952	**.886**	.942	**.878**	**.939**	**.879**	**.940**	**.871**	**.941**	**.874**	**.937**	**.865**	**.930**	**.858**	**.929**	**.854**	.946	**.865**	**.919**	**.857**
	1	.944	**.878**	.944	**.872**	.949	**.877**	.943	**.873**	.942	**.873**	.945	**.862**	**.941**	**.869**	**.938**	**.868**	**.937**	**.860**	**.934**	**.854**	.953	**.869**	**.925**	**.862**
	2	.945	**.881**	**.939**	**.869**	.954	**.891**	.949	**.883**	.946	**.869**	.942	**.864**	.951	**.874**	**.937**	**.872**	**.933**	**.849**	**.933**	**.844**	.941	**.868**	**.926**	**.861**
	4	.946	**.883**	**.940**	**.873**	.953	**.884**	.952	**.881**	**.940**	**.866**	**.936**	**.859**	.949	**.867**	**.939**	**.863**	.941	**.862**	**.938**	**.858**	**.937**	**.859**	**.919**	**.851**
2	-1	.949	**.872**	.944	**.863**	.957	**.885**	.942	**.879**	**.939**	**.871**	**.938**	**.862**	.949	**.872**	**.924**	**.867**	**.936**	**.863**	**.937**	**.855**	**.940**	**.859**	**.922**	**.853**
	0	.947	**.873**	.947	**.866**	.946	**.879**	.947	**.869**	**.936**	**.880**	**.940**	**.873**	.951	**.876**	**.936**	**.873**	**.934**	**.856**	**.931**	**.852**	.944	**.854**	**.925**	**.844**
	1	.956	**.872**	.946	**.863**	.950	**.881**	.944	**.879**	**.938**	**.867**	**.937**	**.861**	.951	**.874**	.942	**.872**	**.934**	**.861**	**.938**	**.858**	.945	**.863**	**.934**	**.861**
	2	.943	**.879**	**.940**	**.867**	.948	**.878**	.949	**.871**	**.933**	**.866**	**.932**	**.860**	.953	**.871**	.942	**.870**	.943	**.862**	.942	**.856**	**.933**	**.864**	**.919**	**.856**
	4	.945	**.872**	.942	**.863**	.948	**.875**	.950	**.871**	.942	**.872**	**.937**	**.865**	**.941**	**.873**	**.937**	**.870**	**.928**	**.861**	**.931**	**.859**	**.938**	**.870**	**.922**	**.866**
3	-1	.950	**.863**	.943	**.853**	.945	**.876**	.944	**.872**	**.938**	**.859**	**.931**	**.854**	**.941**	**.883**	**.933**	**.878**	**.930**	**.861**	**.930**	**.855**	.946	**.874**	**.920**	**.868**
	0	.948	**.876**	.947	**.865**	.951	**.871**	.951	**.866**	.943	**.872**	**.938**	**.866**	**.936**	**.871**	**.922**	**.866**	**.933**	**.864**	**.933**	**.857**	.942	**.858**	**.921**	**.854**
	1	.950	**.889**	.949	**.877**	.949	**.873**	.946	**.868**	**.933**	**.869**	**.937**	**.861**	.949	**.878**	**.935**	**.875**	**.936**	**.863**	**.936**	**.859**	.944	**.861**	**.927**	**.855**
	2	.949	**.886**	**.939**	**.873**	.955	**.886**	.948	**.876**	.947	**.868**	.946	**.860**	**.941**	**.874**	**.934**	**.869**	**.937**	**.852**	**.929**	**.848**	.944	**.859**	**.914**	**.855**
	4	.950	**.880**	.947	**.871**	.956	**.887**	.945	**.886**	**.939**	**.872**	.943	**.865**	.949	**.871**	**.935**	**.865**	**.931**	**.868**	**.933**	**.863**	.943	**.864**	**.930**	**.863**

Note. Skew = skewness of the level-1 error variance, Kurt = kurtosis of the level-1 error variance, ReML = restricted maximum likelihood estimation, MCMC = Markov chain Monte Carlo estimation, W = within-series model, B = between-series model, *I* = number of measurement occasions, *J* = number of participants. Values in Bold are in the unacceptable range [.942 ,.958].

Table 8.
CI Width for the Treatment Effect

		I=10								*I=20*								*I=40*							
		J = 4				*J = 8*				*J = 4*				*J = 8*				*J = 4*				*J = 8*			
		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC	
Skew	Kurt	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B
0	-1	3.69	3.31	3.67	3.22	2.33	2.35	2.54	2.32	2.60	2.94	2.42	2.89	1.64	2.08	1.66	2.06	2.28	2.88	1.91	2.85	1.37	2.04	1.30	2.02
	0	3.65	3.31	3.67	3.22	2.33	2.35	2.53	2.32	2.62	2.94	2.42	2.89	1.64	2.08	1.67	2.06	2.22	2.88	1.91	2.85	1.36	2.04	1.30	2.02
	1	3.66	3.31	3.67	3.22	2.32	2.35	2.54	2.32	2.64	2.94	2.43	2.89	1.65	2.08	1.66	2.06	2.26	2.88	1.91	2.85	1.37	2.04	1.30	2.02
	2	3.63	3.32	3.67	3.22	2.33	2.34	2.53	2.31	2.57	2.93	2.42	2.88	1.64	2.08	1.66	2.06	2.25	2.88	1.91	2.85	1.37	2.04	1.30	2.02
	4	3.64	3.30	3.67	3.21	2.33	2.35	2.53	2.31	2.60	2.93	2.42	2.89	1.64	2.08	1.66	2.06	2.25	2.88	1.91	2.85	1.37	2.04	1.31	2.02
1	-1	3.62	3.30	3.67	3.21	2.32	2.35	2.53	2.31	2.63	2.94	2.42	2.89	1.64	2.08	1.66	2.06	2.24	2.88	1.91	2.85	1.36	2.04	1.30	2.02
	0	3.62	3.29	3.67	3.21	2.33	2.35	2.53	2.31	2.61	2.94	2.42	2.89	1.63	2.08	1.66	2.06	2.25	2.88	1.91	2.85	1.37	2.04	1.30	2.02
	1	3.64	3.30	3.66	3.22	2.32	2.34	2.53	2.31	2.58	2.94	2.42	2.89	1.63	2.08	1.66	2.06	2.26	2.88	1.92	2.85	1.37	2.04	1.30	2.02
	2	3.64	3.31	3.68	3.22	2.33	2.35	2.54	2.32	2.64	2.94	2.43	2.89	1.65	2.08	1.66	2.06	2.24	2.88	1.92	2.85	1.36	2.04	1.29	2.02
	4	3.61	3.30	3.67	3.21	2.33	2.35	2.54	2.32	2.62	2.93	2.42	2.88	1.66	2.08	1.67	2.06	2.25	2.88	1.92	2.85	1.37	2.04	1.30	2.02
2	-1	3.64	3.32	3.68	3.23	2.33	2.35	2.54	2.32	2.61	2.94	2.41	2.89	1.65	2.08	1.66	2.06	2.26	2.88	1.91	2.85	1.36	2.04	1.29	2.02
	0	3.67	3.31	3.67	3.22	2.33	2.35	2.54	2.32	2.58	2.93	2.41	2.88	1.64	2.08	1.66	2.06	2.27	2.88	1.91	2.85	1.38	2.04	1.30	2.02
	1	3.65	3.30	3.67	3.21	2.32	2.35	2.53	2.32	2.63	2.94	2.43	2.89	1.64	2.08	1.67	2.06	2.26	2.88	1.91	2.85	1.36	2.04	1.30	2.02
	2	3.62	3.31	3.67	3.22	2.32	2.35	2.53	2.32	2.64	2.93	2.42	2.89	1.64	2.08	1.66	2.06	2.27	2.88	1.91	2.85	1.37	2.04	1.30	2.02
	4	3.62	3.30	3.66	3.21	2.33	2.35	2.53	2.32	2.60	2.94	2.42	2.89	1.64	2.08	1.66	2.06	2.25	2.88	1.91	2.85	1.36	2.04	1.30	2.02
3	-1	3.68	3.31	3.67	3.21	2.33	2.35	2.54	2.32	2.61	2.94	2.42	2.89	1.65	2.08	1.67	2.06	2.26	2.88	1.92	2.85	1.37	2.04	1.30	2.02
	0	3.60	3.30	3.67	3.21	2.32	2.35	2.54	2.32	2.60	2.94	2.41	2.89	1.64	2.08	1.66	2.06	2.26	2.88	1.92	2.85	1.37	2.04	1.30	2.02
	1	3.61	3.30	3.67	3.21	2.32	2.35	2.53	2.31	2.58	2.94	2.41	2.89	1.64	2.08	1.66	2.06	2.26	2.88	1.91	2.85	1.37	2.04	1.30	2.02
	2	3.63	3.31	3.67	3.22	2.34	2.35	2.53	2.32	2.63	2.93	2.41	2.89	1.63	2.08	1.66	2.06	2.24	2.88	1.90	2.85	1.36	2.04	1.30	2.02
	4	3.67	3.31	3.68	3.22	2.33	2.35	2.53	2.32	2.62	2.93	2.41	2.89	1.65	2.08	1.66	2.06	2.23	2.88	1.91	2.85	1.37	2.04	1.30	2.02

Note. Skew = skewness of the level-1 error variance, Kurt = kurtosis of the level-1 error variance, ReML = restricted maximum likelihood estimation, MCMC = Markov chain Monte Carlo estimation, W = within-series model, B = between-series model, *I* = number of measurement occasions, *J* = number of participants.

Table 9.
Statistical Power for the Treatment Effect

		I=10								*I=20*								*I=40*							
		J = 4				*J = 8*				*J = 4*				*J = 8*				*J = 4*				*J = 8*			
		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC	
Skew	Kurt	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B
0	-1	.208	.299	.200	.313	.376	.388	.326	.395	.339	.315	.355	.323	.665	.473	.659	.479	.443	.343	.528	.348	.796	.487	.817	.493
	0	.185	.268	.189	.284	.384	.404	.338	.413	.365	.319	.381	.330	.675	.482	.655	.492	.460	.338	.537	.342	.811	.508	.834	.508
	1	.189	.275	.197	.288	.397	.417	.341	.424	.331	.324	.364	.329	.667	.476	.650	.483	.442	.325	.525	.331	.805	.485	.830	.490
	2	.186	.285	.189	.302	.390	.407	.333	.419	.365	.318	.369	.332	.658	.484	.640	.489	.461	.329	.547	.336	.804	.491	.839	.498
	4	.189	.284	.199	.297	.380	.411	.329	.423	.352	.313	.377	.321	.666	.486	.652	.495	.450	.339	.554	.346	.821	.490	.828	.493
1	-1	.197	.285	.191	.304	.394	.410	.342	.417	.349	.309	.377	.317	.649	.471	.627	.476	.458	.331	.539	.336	.802	.493	.827	.495
	0	.191	.284	.196	.298	.386	.409	.347	.415	.343	.314	.371	.328	.662	.470	.645	.477	.435	.329	.521	.335	.823	.480	.842	.489
	1	.207	.290	.197	.306	.403	.421	.353	.432	.357	.313	.378	.326	.666	.474	.647	.477	.463	.318	.535	.323	.826	.498	.841	.504
	2	.181	.268	.196	.280	.390	.414	.332	.423	.333	.305	.359	.313	.648	.477	.638	.483	.452	.336	.536	.341	.804	.484	.822	.490
	4	.187	.277	.190	.289	.406	.427	.347	.434	.358	.318	.379	.326	.647	.473	.646	.480	.446	.322	.535	.330	.792	.472	.812	.476
2	-1	.191	.270	.190	.285	.382	.412	.347	.421	.329	.310	.364	.320	.647	.465	.639	.475	.450	.324	.528	.333	.817	.482	.835	.491
	0	.188	.274	.200	.293	.394	.419	.347	.423	.361	.311	.363	.323	.657	.474	.644	.481	.430	.317	.528	.325	.813	.469	.830	.478
	1	.176	.276	.184	.290	.376	.411	.327	.424	.341	.334	.345	.342	.662	.491	.638	.496	.441	.343	.513	.350	.824	.500	.843	.502
	2	.207	.288	.201	.300	.385	.396	.346	.404	.348	.326	.375	.338	.669	.470	.652	.481	.433	.309	.531	.311	.804	.481	.833	.486
	4	.211	.279	.212	.291	.383	.420	.340	.426	.342	.303	.367	.311	.650	.462	.635	.469	.436	.320	.532	.326	.813	.486	.828	.490
3	-1	.199	.282	.192	.298	.402	.403	.358	.413	.353	.328	.367	.337	.658	.465	.644	.471	.441	.330	.533	.337	.819	.493	.829	.497
	0	.201	.280	.187	.300	.387	.402	.320	.411	.340	.319	.370	.331	.677	.478	.660	.485	.443	.351	.529	.361	.806	.500	.832	.500
	1	.192	.265	.193	.281	.409	.419	.346	.426	.363	.317	.391	.324	.656	.458	.650	.464	.458	.305	.537	.310	.813	.482	.830	.488
	2	.192	.280	.190	.292	.393	.419	.345	.425	.336	.312	.362	.322	.666	.482	.647	.490	.447	.334	.534	.343	.801	.496	.827	.501
	4	.188	.267	.202	.284	.383	.414	.326	.425	.338	.322	.376	.331	.660	.487	.644	.490	.452	.323	.540	.331	.800	.482	.825	.490

Note. Skew = skewness of the level-1 error variance, Kurt = kurtosis of the level-1 error variance, ReML = restricted maximum likelihood estimation, MCMC = Markov chain Monte Carlo estimation, W = within-series model, B = between-series model, *I* = number of measurement occasions, *J* = number of participants.

Table 10.

Type I Error for the Treatment Effect

		I=10								*I=20*								*I=40*							
		J = 4				*J = 8*				*J = 4*				*J = 8*				*J = 4*				*J = 8*			
		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC	
Skew	Kurt	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B
0	-1	.050	.120	.055	.131	.047	.121	.054	.125	.051	.132	.065	.141	.058	.121	.071	.126	.064	.141	.068	.144	.059	.137	.082	.140
	0	.052	.128	.044	.138	.043	.115	.055	.117	.054	.123	.062	.128	.049	.122	.061	.123	.065	.139	.070	.144	.063	.149	.088	.148
	1	.050	.120	.065	.130	.051	.122	.051	.127	.059	.134	.065	.137	.055	.142	.066	.144	.064	.147	.067	.151	.060	.136	.079	.141
	2	.046	.112	.051	.121	.050	.133	.057	.137	.051	.121	.068	.128	.051	.122	.064	.128	.062	.131	.074	.135	.059	.129	.078	.130
	4	.054	.123	.055	.133	.055	.135	.055	.140	.057	.140	.063	.144	.057	.125	.074	.127	.066	.141	.065	.148	.056	.128	.076	.133
1	-1	.047	.118	.051	.128	.054	.131	.059	.138	.061	.128	.067	.133	.052	.126	.061	.133	.065	.133	.070	.137	.060	.129	.075	.130
	0	.052	.123	.058	.132	.047	.119	.057	.126	.059	.131	.064	.136	.057	.127	.075	.132	.068	.132	.065	.138	.057	.128	.077	.131
	1	.050	.124	.054	.138	.048	.114	.053	.118	.054	.127	.070	.133	.049	.133	.065	.139	.065	.135	.064	.144	.049	.136	.081	.139
	2	.054	.123	.059	.135	.048	.119	.053	.125	.057	.131	.058	.140	.056	.137	.075	.139	.064	.136	.070	.141	.049	.141	.068	.146
	4	.053	.113	.053	.127	.055	.123	.057	.128	.056	.134	.071	.142	.058	.123	.068	.130	.068	.141	.067	.144	.059	.136	.079	.138
2	-1	.044	.122	.056	.130	.044	.111	.050	.116	.049	.131	.065	.142	.057	.120	.073	.126	.072	.132	.072	.138	.061	.139	.080	.144
	0	.048	.129	.052	.139	.053	.122	.057	.126	.058	.133	.062	.141	.049	.132	.070	.133	.064	.136	.070	.142	.054	.135	.072	.140
	1	.057	.132	.062	.144	.049	.120	.055	.125	.059	.124	.068	.130	.060	.119	.070	.123	.072	.134	.073	.141	.051	.131	.069	.139
	2	.053	.117	.057	.124	.052	.120	.056	.125	.061	.137	.068	.142	.055	.136	.065	.142	.058	.129	.067	.135	.057	.139	.078	.143
	4	.052	.128	.053	.142	.050	.118	.053	.122	.049	.140	.063	.143	.051	.118	.065	.122	.063	.141	.060	.147	.059	.150	.072	.156
3	-1	.050	.129	.051	.138	.046	.115	.050	.123	.050	.122	.054	.127	.049	.124	.068	.129	.065	.126	.066	.130	.061	.145	.082	.145
	0	.045	.126	.053	.135	.056	.125	.054	.129	.059	.128	.070	.134	.055	.118	.074	.123	.064	.133	.066	.136	.058	.142	.079	.144
	1	.054	.130	.058	.141	.052	.123	.054	.128	.051	.133	.064	.142	.059	.124	.071	.129	.061	.132	.067	.139	.059	.129	.081	.132
	2	.055	.120	.055	.126	.045	.123	.049	.129	.062	.133	.060	.144	.052	.128	.072	.133	.056	.146	.062	.149	.060	.144	.070	.149
	4	.055	.128	.058	.138	.053	.123	.059	.128	.054	.124	.065	.129	.061	.138	.078	.138	.071	.149	.074	.154	.060	.136	.083	.139

Note. Skew = skewness of the level-1 error variance, Kurt = kurtosis of the level-1 error variance, ReML = restricted maximum likelihood estimation, MCMC = Markov chain Monte Carlo estimation, W = within-series model, B = between-series model, *I* = number of measurement occasions, *J* = number of participants.

Table 11.

The Treatment Effect Parameter Estimate

		I=10								*I=20*								*I=40*							
		J = 4				*J = 8*				*J = 4*				*J = 8*				*J = 4*				*J = 8*			
		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC	
Skew	Kurt	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B
0	-1	1.03	1.02	1.02	1.02	.99	.99	.98	.99	.99	1.00	.99	.99	1.01	1.00	1.01	1.00	1.00	1.01	1.00	1.01	.99	.97	.99	.97
	0	.98	.96	.98	.96	1.00	1.00	1.00	1.00	1.01	1.02	1.01	1.02	1.00	1.01	1.00	1.01	1.00	1.01	.99	1.01	1.01	1.02	1.00	1.02
	1	1.00	1.00	1.02	1.00	1.00	1.00	1.01	1.00	1.00	1.01	1.00	1.01	1.01	1.00	1.00	1.00	1.00	.98	.99	.98	1.01	.99	1.01	.99
	2	1.00	1.01	.99	1.01	1.00	.99	1.00	.99	1.00	1.02	1.00	1.02	.99	.99	.99	.99	1.01	.99	1.00	.99	.99	.99	1.00	.99
	4	1.01	1.00	.99	1.00	.99	1.00	.99	1.00	1.00	.99	1.00	.99	1.00	1.02	1.00	1.02	1.01	1.01	1.00	1.01	1.00	1.00	1.00	1.00
1	-1	1.00	1.00	1.00	1.00	1.00	1.00	.99	1.00	1.00	1.01	1.00	1.01	.99	.99	.98	.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0	1.01	1.00	1.01	1.00	1.00	1.00	1.00	1.00	.99	1.01	.99	1.01	1.00	.99	1.00	.99	.98	1.03	.98	1.02	1.01	1.00	1.01	1.00
	1	1.02	1.04	1.02	1.04	1.02	1.02	1.01	1.02	1.01	1.00	1.01	1.00	1.00	.99	1.00	.99	1.01	.99	1.01	.99	1.01	1.02	1.01	1.02
	2	.98	.97	.99	.97	1.00	1.02	1.00	1.02	.99	.97	.98	.97	.99	1.00	.99	1.00	1.01	1.01	1.00	1.01	.98	1.00	.98	1.00
	4	1.00	1.01	1.00	1.01	1.02	1.01	1.01	1.01	1.02	1.03	1.01	1.03	1.00	1.02	1.00	1.01	.99	.99	.99	.99	.99	.99	.99	.99
2	-1	1.00	1.00	.99	1.00	1.00	1.01	1.00	1.01	.98	.98	.98	.98	.99	.98	.99	.98	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.01
	0	1.00	.99	1.00	.99	1.01	1.01	1.01	1.01	1.00	1.01	1.00	1.01	1.00	1.00	.99	1.00	.99	.99	.99	.99	1.00	.99	1.00	.99
	1	.96	.97	.98	.97	.98	1.01	.99	1.01	1.00	1.02	1.00	1.02	1.00	1.00	1.00	1.00	.99	1.01	.98	1.01	1.01	1.01	1.01	1.01
	2	1.01	.99	1.00	1.00	1.00	.99	1.00	.99	1.02	1.03	1.01	1.03	1.00	1.01	1.00	1.01	.99	.98	.98	.98	1.00	.99	1.00	.99
	4	1.03	.99	1.03	.99	1.01	1.01	1.00	1.01	.99	.98	.98	.98	.99	.99	.99	.99	.99	.99	.99	.99	1.00	1.00	1.00	1.00
3	-1	1.01	1.01	.99	1.01	1.02	1.01	1.02	1.01	1.00	1.02	1.00	1.02	.99	.98	1.00	.98	1.01	1.02	1.01	1.02	1.00	1.00	1.00	1.00
	0	1.01	1.02	1.02	1.02	.99	.98	.99	.98	.98	.99	.99	.99	1.02	.99	1.02	.99	1.00	1.02	1.00	1.02	1.01	1.00	1.00	1.00
	1	1.02	.99	1.01	.99	1.01	1.01	1.00	1.01	1.01	1.03	1.01	1.03	1.00	1.00	.99	1.00	.99	.97	.99	.97	1.00	.99	1.00	.99
	2	1.00	1.00	.99	1.00	1.01	1.01	1.00	1.01	.99	.98	.99	.98	1.00	1.00	1.01	1.00	1.00	1.01	.99	1.01	1.00	1.00	1.00	1.00
	4	1.00	.99	1.01	.99	.99	1.01	.98	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.01	.99	.98	.99	.98

Note. Skew = skewness of the level-1 error variance, Kurt = kurtosis of the level-1 error variance, ReML = restricted maximum likelihood estimation, MCMC = Markov chain Monte Carlo estimation, W = within-series model, B = between-series model, *I* = number of measurement occasions, *J* = number of participants.

Table 12.
Bias of Intercept Parameter Estimate in the Within-Series Model

		I=10				*I=20*				*I=40*			
		J = 4		*J = 8*		*J = 4*		*J = 8*		*J = 4*		*J = 8*	
Skew	Kurt	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC
0	-1	.001	.001	-.006	-.007	.006	.005	-.006	-.003	-.004	-.003	.003	.005
	0	-.009	-.009	-.004	-.005	-.005	-.003	-.004	-.006	-.001	-.005	-.004	-.007
	1	-.006	-.004	-.005	-.004	-.004	-.005	.000	-.001	-.003	-.005	.002	.002
	2	.006	.005	-.005	-.005	-.002	-.004	.003	.000	-.002	-.003	-.002	-.001
	4	.007	.003	-.004	-.003	.007	.008	.002	.002	-.007	-.008	.002	.002
1	-1	-.002	-.001	-.002	-.002	-.014	-.014	-.004	-.005	-.010	-.010	.009	.011
	0	.005	.003	.006	.007	.000	-.002	.003	.002	-.003	-.007	-.008	-.009
	1	.008	.006	-.003	-.006	-.013	-.014	-.002	-.002	-.015	-.019	.007	.006
	2	-.016	-.011	.010	.008	-.007	-.008	-.004	-.004	-.007	-.010	.009	.007
	4	.005	.005	.002	.002	-.003	-.004	-.002	.000	.007	.008	.003	.003
2	-1	-.012	-.012	-.003	-.002	-.006	-.003	.002	.002	.011	.010	-.007	-.008
	0	.000	-.003	.005	.003	.008	.005	.011	.009	.004	.004	.002	.000
	1	-.011	-.006	-.018	-.017	-.012	-.009	-.004	-.002	.009	.006	-.001	.000
	2	.005	.001	-.001	.000	.012	.009	.003	.003	-.006	-.008	.002	.004
	4	.009	.009	-.001	-.003	.012	.010	.001	.000	.008	.007	.001	.002
3	-1	-.006	-.008	.004	.006	.000	-.004	-.008	-.006	.005	.005	-.001	-.002
	0	-.006	-.004	.004	.006	-.004	-.003	.008	.010	.014	.012	.000	-.002
	1	.005	.004	.011	.009	.006	.004	-.004	-.007	.002	.007	.003	.001
	2	-.011	-.012	-.007	-.008	-.008	-.006	.001	.003	-.004	-.007	.000	.000
	4	.004	.007	-.006	-.009	.007	.006	.000	-.001	-.011	-.012	-.001	.001

Note. Skew = skewness of the level-1 error variance, Kurt = kurtosis of the level-1 error variance, ReML = restricted maximum likelihood estimation, MCMC = Markov chain Monte Carlo estimation, W = within-series model, B = between-series model, *I* = number of measurement occasions, *J* = number of participants.

Table 13.
RMSE of Intercept Parameter Estimate in the Within-Series Model

		I=10				*I=20*				*I=40*			
		J = 4		*J = 8*		*J = 4*		*J = 8*		*J = 4*		*J = 8*	
Skew	Kurt	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC
0	-1	.481	.493	.330	.336	.427	.447	.303	.315	.395	.412	.274	.285
	0	.493	.504	.342	.354	.424	.444	.303	.316	.394	.410	.283	.297
	1	.460	.473	.335	.346	.427	.443	.303	.315	.394	.410	.276	.290
	2	.484	.495	.341	.351	.427	.446	.308	.318	.404	.421	.288	.298
	4	.475	.486	.336	.346	.427	.444	.307	.319	.387	.406	.279	.286
1	-1	.469	.482	.336	.344	.440	.460	.305	.321	.399	.415	.279	.290
	0	.491	.506	.336	.348	.439	.455	.302	.315	.407	.422	.285	.295
	1	.484	.497	.341	.349	.425	.445	.300	.315	.400	.416	.293	.302
	2	.469	.486	.338	.348	.435	.454	.300	.311	.407	.428	.277	.290
	4	.485	.499	.338	.348	.438	.454	.302	.310	.400	.416	.283	.293
2	-1	.477	.491	.329	.339	.442	.456	.307	.321	.395	.412	.281	.293
	0	.479	.491	.335	.342	.431	.446	.302	.315	.402	.422	.288	.300
	1	.473	.491	.333	.345	.437	.454	.303	.315	.397	.417	.276	.290
	2	.479	.498	.335	.344	.438	.457	.298	.310	.389	.405	.281	.292
	4	.486	.494	.339	.348	.434	.451	.302	.311	.400	.417	.279	.290
3	-1	.482	.495	.338	.349	.431	.449	.308	.321	.399	.412	.285	.295
	0	.483	.499	.332	.339	.444	.459	.303	.316	.397	.415	.281	.293
	1	.480	.490	.342	.349	.437	.455	.310	.319	.404	.421	.283	.295
	2	.473	.485	.341	.348	.428	.445	.300	.313	.396	.412	.281	.293
	4	.476	.491	.335	.344	.445	.460	.307	.319	.397	.412	.285	.295

Note. Skew = skewness of the level-1 error variance, Kurt = kurtosis of the level-1 error variance, ReML = restricted maximum likelihood estimation, MCMC = Markov chain Monte Carlo estimation, W = within-series model, B = between-series model, *I* = number of measurement occasions, *J* = number of participants.

Table 14.
Bias of Time Effect in the Baseline Phase in the Within-Series Model

		I=10				*I=20*				*I=40*			
		J = 4		*J = 8*		*J = 4*		*J = 8*		*J = 4*		*J = 8*	
Skew	Kurt	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC
0	-1	-.002	-.002	.003	.004	.001	.002	.000	-.001	.000	.000	.000	.000
	0	.002	.001	.002	.002	.000	-.001	.000	.001	.000	.001	.000	.000
	1	-.002	-.004	.000	-.001	.000	.001	.000	.001	.000	.001	.000	.000
	2	-.001	.000	.001	.000	-.001	-.001	.001	.001	.000	.000	.000	.000
	4	-.002	.001	.001	.001	.000	.000	.000	-.001	.000	.000	.000	.000
1	-1	-.001	-.002	.000	.001	.001	.001	.001	.001	.000	.000	.000	-.001
	0	-.004	-.002	.001	.001	.000	.001	.000	.000	.000	.001	.000	.000
	1	-.003	-.002	-.003	-.001	.000	.000	.000	.000	.000	.000	.000	.000
	2	.004	.001	.000	.001	.000	.001	.000	.000	.000	.000	.000	.000
	4	-.001	-.001	-.001	-.001	-.001	.000	.001	.000	.001	.000	.000	.000
2	-1	.003	.003	.000	.000	.001	.001	.001	.001	.000	.000	.000	.000
	0	-.002	-.001	-.003	-.001	-.001	.000	-.001	.000	.000	.000	.000	.000
	1	.006	.003	.005	.004	.001	-.001	.001	.001	.000	.001	.000	.000
	2	-.003	-.001	.000	-.001	-.002	.000	.000	.000	.000	.001	.000	.000
	4	-.004	-.004	-.003	-.001	.000	.001	.000	.000	.000	.000	.000	.000
3	-1	-.001	.001	-.002	-.003	.000	.001	.000	.000	.000	.000	.000	.000
	0	.001	.000	.001	.000	.001	.000	-.001	-.002	.000	.000	.000	.000
	1	-.004	-.003	-.002	.000	.000	.001	.000	.001	.001	.000	.000	.000
	2	.001	.001	.000	.000	.000	-.001	.000	-.001	.000	.001	.000	.000
	4	-.003	-.005	.001	.003	.000	.000	-.001	.000	.000	.001	.000	.000

Note. Skew = skewness of the level-1 error variance, Kurt = kurtosis of the level-1 error variance, ReML = restricted maximum likelihood estimation, MCMC = Markov chain Monte Carlo estimation, W = within-series model, B = between-series model, *I* = number of measurement occasions, *J* = number of participants.

Table 15.
RMSE of Time Effect in the Baseline Phase in the Within-Series Model

		I=10				*I=20*				*I=40*			
		J = 4		*J = 8*		*J = 4*		*J = 8*		*J = 4*		*J = 8*	
Skew	Kurt	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC
0	-1	.141	.164	.095	.110	.055	.071	.032	.045	.000	.032	.000	.000
	0	.138	.161	.095	.110	.055	.071	.032	.045	.000	.032	.000	.000
	1	.138	.161	.095	.110	.055	.071	.032	.045	.000	.032	.000	.000
	2	.138	.161	.095	.110	.055	.071	.032	.045	.000	.032	.000	.000
	4	.138	.161	.095	.110	.055	.071	.032	.045	.000	.032	.000	.000
1	-1	.138	.161	.095	.110	.055	.071	.032	.045	.000	.032	.000	.000
	0	.138	.161	.095	.114	.055	.071	.032	.045	.000	.032	.000	.000
	1	.138	.161	.095	.110	.055	.071	.032	.045	.000	.032	.000	.000
	2	.138	.161	.095	.110	.055	.071	.032	.045	.000	.032	.000	.000
	4	.141	.164	.095	.110	.055	.071	.032	.045	.000	.032	.000	.000
2	-1	.134	.161	.089	.110	.055	.071	.032	.045	.000	.032	.000	.000
	0	.138	.161	.095	.110	.055	.071	.032	.045	.000	.032	.000	.000
	1	.134	.161	.095	.110	.055	.071	.032	.045	.000	.032	.000	.000
	2	.138	.161	.095	.114	.055	.071	.032	.045	.000	.032	.000	.000
	4	.138	.161	.095	.110	.055	.071	.032	.045	.000	.032	.000	.000
3	-1	.138	.161	.095	.110	.055	.071	.032	.045	.000	.032	.000	.000
	0	.138	.161	.095	.110	.055	.071	.032	.045	.000	.032	.000	.000
	1	.134	.158	.095	.110	.055	.071	.032	.045	.000	.032	.000	.000
	2	.138	.161	.095	.110	.045	.071	.032	.045	.000	.032	.000	.000
	4	.138	.164	.095	.110	.055	.071	.032	.045	.000	.032	.000	.000

Note. Skew = skewness of the level-1 error variance, Kurt = kurtosis of the level-1 error variance, ReML = restricted maximum likelihood estimation, MCMC = Markov chain Monte Carlo estimation, W = within-series model, B = between-series model, *I* = number of measurement occasions, *J* = number of participants.

Table 16.
Bias of Time Effect in the Treatment Phase in the Within-Series Model

		I=10				*I=20*				*I=40*			
		J = 4		*J = 8*		*J = 4*		*J = 8*		*J = 4*		*J = 8*	
Skew	Kurt	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC
0	-1	.003	.002	-.001	-.001	-.001	-.001	.000	.001	.000	.000	.000	.000
	0	-.001	-.001	-.002	-.002	.000	.001	.000	.000	.000	.000	.000	.000
	1	.005	.009	.005	.005	.000	.000	-.001	-.002	.000	-.001	.000	.000
	2	-.001	-.004	-.001	-.001	.001	.001	-.001	-.003	.000	.000	.000	.000
	4	-.005	-.012	-.002	-.003	.000	-.001	.001	.001	.000	.000	.000	.000
1	-1	.003	.003	-.001	-.003	-.002	-.002	-.001	.000	.000	.000	.000	.000
	0	.003	.005	-.002	-.001	.000	-.002	.001	.001	.000	-.001	.000	.000
	1	.000	.001	.004	.002	-.001	-.001	.000	.001	.000	-.001	.000	.000
	2	-.004	.001	.001	-.003	.000	-.001	.000	.000	.000	.000	.000	-.001
	4	-.003	-.002	.001	.000	-.001	-.002	.000	.000	.000	.000	.000	.000
2	-1	-.002	-.003	-.001	.000	-.001	-.002	.000	.000	.000	.000	.000	.000
	0	.003	.002	.003	.002	.001	.000	.001	.000	.000	.000	.000	.000
	1	-.001	.001	-.002	-.001	-.001	.000	-.002	.000	-.001	.000	.000	.000
	2	.001	.002	.001	.002	.001	-.001	.000	-.001	-.001	-.001	.000	.000
	4	.004	.007	.002	.000	.001	-.002	.000	-.001	.000	-.001	.000	.000
3	-1	-.001	-.004	.002	.003	.000	-.002	-.001	.001	.000	.000	.000	.000
	0	.003	.002	-.002	-.001	-.001	.001	.000	.002	-.001	-.001	.000	.000
	1	.001	.001	.001	-.001	.001	.000	-.001	-.003	.000	.000	.000	-.001
	2	.002	-.002	-.001	-.004	.001	.001	.001	.002	.000	.000	.000	.000
	4	.003	.007	-.002	-.006	.000	-.001	.000	.000	.000	.000	.000	.000

Note. Skew = skewness of the level-1 error variance, Kurt = kurtosis of the level-1 error variance, ReML = restricted maximum likelihood estimation, MCMC = Markov chain Monte Carlo estimation, W = within-series model, B = between-series model, *I* = number of measurement occasions, *J* = number of participants.

Table 17.
RMSE of Time Effect in the Treatment Phase in the Within-Series Model

		I=10				*I=20*				*I=40*			
		J = 4		*J = 8*		*J = 4*		*J = 8*		*J = 4*		*J = 8*	
Skew	Kurt	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC
0	-1	.205	.261	.134	.173	.071	.100	.045	.071	.032	.032	.000	.032
	0	.210	.263	.134	.173	.071	.100	.045	.071	.032	.032	.000	.032
	1	.205	.257	.130	.173	.071	.100	.045	.071	.032	.032	.000	.032
	2	.200	.253	.130	.173	.071	.100	.045	.071	.032	.032	.000	.032
	4	.205	.259	.134	.173	.071	.100	.045	.071	.032	.032	.000	.032
1	-1	.207	.257	.138	.176	.071	.100	.045	.071	.032	.032	.000	.032
	0	.200	.253	.134	.179	.071	.100	.045	.071	.032	.032	.000	.032
	1	.207	.259	.134	.176	.071	.095	.045	.071	.032	.032	.000	.032
	2	.205	.257	.134	.173	.071	.100	.045	.071	.032	.032	.000	.032
	4	.205	.259	.134	.173	.071	.100	.045	.071	.032	.032	.000	.032
2	-1	.200	.261	.130	.173	.071	.100	.045	.071	.032	.032	.000	.032
	0	.202	.253	.141	.179	.071	.100	.045	.071	.032	.032	.000	.032
	1	.207	.261	.134	.176	.071	.100	.045	.071	.032	.032	.000	.032
	2	.207	.259	.138	.179	.071	.100	.045	.071	.032	.032	.000	.032
	4	.200	.253	.134	.176	.071	.100	.045	.071	.032	.032	.000	.032
3	-1	.207	.257	.134	.176	.071	.100	.045	.071	.032	.032	.000	.032
	0	.207	.259	.134	.173	.071	.100	.045	.071	.032	.032	.000	.032
	1	.202	.257	.138	.176	.071	.100	.045	.071	.032	.032	.000	.032
	2	.202	.255	.134	.173	.071	.100	.045	.071	.032	.032	.000	.032
	4	.200	.257	.134	.176	.071	.100	.045	.071	.032	.032	.000	.032

Note. Skew = skewness of the level-1 error variance, Kurt = kurtosis of the level-1 error variance, ReML = restricted maximum likelihood estimation, MCMC = Markov chain Monte Carlo estimation, W = within-series model, B = between-series model, *I* = number of measurement occasions, *J* = number of participants.

Table 18.
Relative Bias of Level-1 Error Variance Estimate

		I=10								*I=20*								*I=40*							
		J = 4				*J = 8*				*J = 4*				*J = 8*				*J = 4*				*J = 8*			
		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC	
Skew	Kurt	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B
0	-1	-.049	.103	-.017	.134	-.025	.155	-.017	.177	-.017	.041	.003	.053	-.010	.060	.002	.070	-.008	.020	.004	.024	-.006	.029	.004	.033
	0	-.054	.105	-.022	.137	-.028	.156	-.019	.178	-.018	.039	.001	.050	-.010	.059	.003	.070	-.008	.019	.004	.023	-.005	.029	.005	.033
	1	-.050	.101	-.016	.133	-.023	.157	-.015	.177	-.017	.038	.003	.051	-.010	.059	.003	.070	-.007	.020	.005	.024	-.006	.029	.004	.033
	2	-.051	.107	-.020	.137	-.027	.150	-.020	.171	-.019	.037	.000	.048	-.010	.058	.002	.069	-.008	.020	.004	.024	-.005	.030	.006	.034
	4	-.051	.098	-.019	.131	-.027	.154	-.018	.176	-.021	.037	-.001	.049	-.010	.059	.003	.069	-.008	.021	.005	.025	-.005	.029	.005	.033
1	-1	-.056	.096	-.021	.127	-.027	.156	-.018	.177	-.020	.040	.001	.051	-.011	.059	.001	.069	-.008	.019	.005	.023	-.005	.030	.005	.034
	0	-.052	.092	-.022	.124	-.026	.153	-.018	.176	-.017	.039	.003	.051	-.009	.061	.003	.071	-.007	.019	.005	.023	-.005	.030	.005	.034
	1	-.056	.099	-.025	.132	-.027	.151	-.021	.173	-.017	.040	.003	.051	-.010	.061	.003	.071	-.008	.019	.004	.023	-.005	.029	.004	.033
	2	-.049	.099	-.018	.132	-.024	.160	-.016	.181	-.018	.039	.004	.051	-.011	.060	.002	.071	-.008	.019	.004	.024	-.005	.030	.005	.034
	4	-.051	.098	-.020	.129	-.026	.161	-.017	.183	-.020	.036	.000	.047	-.010	.062	.003	.073	-.008	.020	.004	.024	-.005	.029	.004	.033
2	-1	-.049	.109	-.017	.139	-.025	.156	-.017	.178	-.019	.041	.001	.052	-.010	.061	.003	.072	-.008	.020	.005	.024	-.005	.029	.005	.033
	0	-.049	.103	-.017	.134	-.026	.158	-.017	.181	-.018	.037	.002	.048	-.011	.059	.003	.070	-.008	.019	.004	.023	-.005	.030	.005	.034
	1	-.053	.096	-.021	.127	-.025	.155	-.015	.177	-.018	.040	.003	.051	-.011	.060	.001	.070	-.008	.019	.004	.023	-.005	.030	.005	.034
	2	-.052	.105	-.020	.135	-.028	.158	-.020	.180	-.019	.038	.001	.050	-.010	.059	.002	.069	-.007	.020	.005	.024	-.005	.029	.004	.034
	4	-.055	.095	-.025	.127	-.029	.152	-.021	.175	-.017	.040	.004	.052	-.011	.060	.003	.071	-.008	.020	.005	.024	-.005	.030	.005	.034
3	-1	-.055	.100	-.024	.130	-.027	.158	-.018	.180	-.019	.039	.001	.051	-.010	.060	.002	.071	-.008	.020	.004	.024	-.005	.029	.004	.033
	0	-.051	.097	-.019	.128	-.025	.159	-.017	.180	-.018	.039	.002	.051	-.012	.059	.001	.070	-.008	.021	.005	.026	-.005	.029	.005	.033
	1	-.049	.098	-.020	.130	-.026	.154	-.019	.175	-.018	.040	.002	.051	-.011	.061	.001	.072	-.007	.019	.006	.024	-.005	.029	.005	.033
	2	-.052	.102	-.019	.133	-.025	.157	-.016	.179	-.018	.037	.002	.049	-.010	.061	.002	.071	-.007	.020	.005	.024	-.005	.028	.005	.032
	4	-.045	.105	-.015	.136	-.025	.157	-.017	.179	-.019	.036	.002	.048	-.010	.061	.002	.071	-.007	.020	.005	.024	-.005	.028	.005	.032

Note. Skew = skewness of the level-1 error variance, Kurt = kurtosis of the level-1 error variance, ReML = restricted maximum likelihood estimation, MCMC = Markov chain Monte Carlo estimation, W = within-series model, B = between-series model, *I* = number of measurement occasions, *J* = number of participants.

Table 19.
RMSE of Level-1 Error Variance Estimate

		I=10								*I=20*								*I=40*							
		J = 4				*J = 8*				*J = 4*				*J = 8*				*J = 4*				*J = 8*			
		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC		ReML		MCMC	
Skew	Kurt	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B	W	B
0	-1	.122	.224	.110	.241	.071	.214	.077	.235	.055	.089	.063	.095	.032	.084	.045	.095	.032	.045	.032	.045	.000	.045	.000	.045
	0	.122	.228	.114	.247	.077	.219	.077	.237	.055	.089	.063	.095	.032	.084	.045	.095	.032	.045	.032	.045	.000	.045	.032	.045
	1	.118	.228	.110	.245	.071	.217	.077	.235	.055	.095	.055	.100	.032	.084	.045	.095	.032	.045	.032	.045	.000	.045	.000	.045
	2	.118	.230	.114	.247	.077	.210	.077	.228	.055	.089	.063	.095	.032	.084	.045	.095	.032	.045	.032	.045	.000	.045	.032	.045
	4	.122	.224	.114	.243	.071	.217	.077	.237	.055	.089	.063	.095	.032	.084	.045	.095	.032	.045	.032	.045	.000	.045	.032	.045
1	-1	.118	.226	.114	.243	.071	.214	.077	.232	.055	.100	.063	.105	.032	.084	.045	.095	.032	.055	.032	.055	.000	.045	.000	.045
	0	.122	.221	.114	.237	.077	.212	.077	.232	.055	.095	.063	.100	.032	.084	.045	.095	.032	.045	.032	.045	.000	.045	.032	.045
	1	.122	.230	.114	.249	.071	.207	.077	.226	.055	.089	.063	.100	.032	.089	.045	.095	.032	.045	.032	.045	.000	.045	.000	.045
	2	.118	.221	.114	.239	.071	.219	.077	.237	.055	.095	.063	.100	.032	.084	.045	.095	.032	.045	.032	.045	.000	.045	.032	.045
	4	.122	.214	.114	.232	.071	.219	.077	.239	.055	.089	.055	.095	.032	.089	.045	.095	.032	.045	.032	.045	.000	.045	.000	.045
2	-1	.118	.228	.114	.247	.071	.217	.077	.235	.055	.095	.063	.100	.032	.089	.045	.095	.032	.045	.032	.045	.000	.045	.000	.045
	0	.118	.224	.114	.241	.071	.221	.077	.241	.055	.089	.055	.095	.032	.084	.045	.095	.032	.045	.032	.045	.000	.045	.000	.045
	1	.118	.224	.114	.241	.071	.217	.077	.235	.055	.095	.063	.100	.032	.084	.045	.095	.032	.045	.032	.055	.000	.045	.032	.045
	2	.118	.226	.114	.243	.077	.217	.077	.237	.055	.095	.055	.105	.032	.084	.045	.095	.032	.045	.032	.045	.000	.045	.032	.045
	4	.122	.221	.114	.239	.077	.212	.077	.232	.055	.095	.055	.100	.032	.084	.045	.095	.032	.045	.032	.045	.000	.045	.032	.045
3	-1	.122	.228	.114	.245	.071	.219	.077	.239	.055	.095	.063	.100	.032	.084	.045	.095	.032	.045	.032	.045	.000	.045	.000	.045
	0	.118	.226	.114	.243	.071	.221	.077	.241	.055	.095	.055	.100	.032	.084	.045	.095	.032	.045	.032	.055	.000	.045	.032	.045
	1	.122	.221	.114	.239	.071	.212	.077	.232	.055	.095	.055	.100	.032	.089	.045	.095	.032	.045	.032	.045	.000	.045	.032	.045
	2	.122	.226	.114	.243	.071	.219	.077	.239	.055	.089	.063	.095	.032	.089	.045	.095	.032	.045	.032	.045	.000	.045	.032	.045
	4	.118	.228	.110	.247	.071	.217	.077	.237	.055	.089	.063	.095	.032	.084	.045	.095	.032	.045	.032	.045	.000	.045	.032	.045

Note. Skew = skewness of the level-1 error variance, Kurt = kurtosis of the level-1 error variance, ReML = restricted maximum likelihood estimation, MCMC = Markov chain Monte Carlo estimation, W = within-series model, B = between-series model, *I* = number of measurement occasions, *J* = number of participants.

Table 20.
Relative Bias of Level-2 Error Variance for the Intercept in Within-Series Model

		I=10				*I=20*				*I=40*			
		J = 4		*J = 8*		*J = 4*		*J = 8*		*J = 4*		*J = 8*	
Skew	Kurt	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC
0	-1	-.250	.060	-.262	-.114	-.116	.086	-.132	-.088	-.076	.102	-.080	-.042
	0	-.226	.060	-.262	-.106	-.144	.066	-.130	-.084	-.064	.114	-.092	-.054
	1	-.220	.068	-.270	-.124	-.122	.086	-.154	-.102	-.082	.106	-.064	-.034
	2	-.266	.052	-.248	-.100	-.134	.076	-.142	-.090	-.070	.106	-.050	-.018
	4	-.220	.074	-.266	-.122	-.158	.070	-.138	-.092	-.118	.078	-.054	-.032
1	-1	-.228	.066	-.258	-.120	-.132	.068	-.148	-.084	-.082	.106	-.058	-.032
	0	-.238	.070	-.246	-.118	-.148	.068	-.126	-.082	-.090	.106	-.070	-.036
	1	-.232	.076	-.270	-.114	-.130	.076	-.132	-.086	-.080	.108	-.090	-.054
	2	-.228	.066	-.248	-.108	-.130	.078	-.144	-.090	-.084	.102	-.064	-.030
	4	-.236	.066	-.238	-.114	-.138	.070	-.118	-.074	-.074	.096	-.060	-.026
2	-1	-.236	.054	-.262	-.126	-.118	.080	-.130	-.078	-.082	.098	-.084	-.042
	0	-.264	.052	-.248	-.114	-.158	.062	-.122	-.080	-.086	.098	-.080	-.044
	1	-.248	.060	-.244	-.114	-.144	.070	-.124	-.082	-.076	.108	-.046	-.026
	2	-.238	.058	-.248	-.112	-.124	.086	-.154	-.084	-.102	.096	-.080	-.042
	4	-.242	.072	-.258	-.114	-.130	.078	-.108	-.074	-.068	.106	-.068	-.038
3	-1	-.236	.072	-.242	-.106	-.134	.068	-.142	-.092	-.064	.122	-.080	-.046
	0	-.230	.056	-.246	-.106	-.172	.066	-.124	-.076	-.072	.106	-.078	-.044
	1	-.230	.076	-.258	-.112	-.126	.078	-.118	-.078	-.080	.092	-.068	-.038
	2	-.246	.052	-.264	-.132	-.128	.072	-.120	-.084	-.098	.102	-.080	-.038
	4	-.232	.064	-.258	-.112	-.110	.086	-.116	-.072	-.092	.094	-.074	-.032

Note. Skew = skewness of the level-1 error variance, Kurt = kurtosis of the level-1 error variance, ReML = restricted maximum likelihood estimation, MCMC = Markov chain Monte Carlo estimation, W = within-series model, B = between-series model, *I* = number of measurement occasions, *J* = number of participants.

Table 21.
RMSE of Level-2 Error Variance for the Intercept in the Within-Series Model

		I=10				*I=20*				*I=40*			
		J = 4		*J = 8*		*J = 4*		*J = 8*		*J = 4*		*J = 8*	
Skew	Kurt	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC
0	-1	.479	.205	.335	.192	.479	.261	.311	.212	.440	.265	.285	.217
	0	.495	.207	.335	.197	.457	.235	.308	.210	.460	.274	.286	.221
	1	.498	.224	.332	.187	.480	.247	.303	.202	.439	.265	.297	.226
	2	.477	.197	.330	.195	.476	.249	.307	.207	.445	.261	.290	.235
	4	.497	.214	.330	.184	.458	.232	.311	.214	.432	.257	.293	.230
1	-1	.500	.212	.330	.192	.466	.226	.307	.219	.444	.274	.285	.226
	0	.485	.212	.332	.190	.470	.224	.310	.207	.444	.274	.293	.224
	1	.500	.230	.330	.190	.482	.243	.311	.219	.449	.272	.283	.219
	2	.500	.214	.333	.200	.471	.239	.303	.205	.437	.268	.283	.221
	4	.483	.210	.324	.182	.475	.247	.313	.210	.449	.261	.293	.237
2	-1	.482	.200	.329	.190	.484	.243	.310	.210	.447	.259	.283	.224
	0	.477	.207	.332	.190	.479	.241	.308	.214	.446	.261	.285	.221
	1	.475	.205	.330	.187	.477	.239	.308	.217	.449	.277	.288	.217
	2	.496	.210	.338	.190	.475	.255	.302	.210	.439	.265	.283	.224
	4	.499	.210	.335	.190	.474	.249	.313	.212	.438	.263	.286	.228
3	-1	.490	.212	.344	.205	.465	.226	.303	.202	.438	.276	.283	.226
	0	.493	.195	.341	.200	.458	.232	.307	.210	.436	.263	.290	.224
	1	.503	.232	.330	.190	.479	.239	.310	.210	.438	.249	.281	.219
	2	.493	.200	.335	.182	.480	.241	.311	.205	.435	.266	.281	.224
	4	.486	.207	.329	.192	.503	.266	.305	.217	.442	.261	.288	.230

Note. Skew = skewness of the level-1 error variance, Kurt = kurtosis of the level-1 error variance, ReML = restricted maximum likelihood estimation, MCMC = Markov chain Monte Carlo estimation, W = within-series model, B = between-series model, *I* = number of measurement occasions, *J* = number of participants.

Table 22.

Relative Bias of Level-2 Error Variance for the Treatment Effect in the Within-Series Model

		I=10				*I=20*				*I=40*			
		J = 4		*J = 8*		*J = 4*		*J = 8*		*J = 4*		*J = 8*	
Skew	Kurt	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC
0	-1	-.030	.078	-.134	-.112	-.108	.048	-.112	-.120	-.076	.056	-.062	-.084
	0	-.084	.078	-.124	-.114	-.100	.048	-.096	-.112	-.116	.058	-.080	-.092
	1	-.096	.074	-.160	-.110	-.058	.062	-.080	-.118	-.086	.052	-.068	-.078
	2	-.072	.070	-.140	-.104	-.130	.052	-.112	-.110	-.102	.056	-.070	-.094
	4	-.092	.076	-.144	-.114	-.100	.042	-.114	-.130	-.100	.056	-.062	-.078
1	-1	-.108	.082	-.158	-.112	-.114	.056	-.094	-.116	-.104	.070	-.082	-.086
	0	-.106	.078	-.156	-.116	-.090	.056	-.128	-.126	-.072	.072	-.074	-.092
	1	-.082	.076	-.150	-.110	-.126	.050	-.128	-.130	-.092	.076	-.078	-.090
	2	-.086	.074	-.118	-.102	-.078	.058	-.078	-.116	-.094	.062	-.090	-.104
	4	-.138	.076	-.142	-.112	-.084	.054	-.062	-.110	-.096	.078	-.074	-.096
2	-1	-.094	.072	-.128	-.116	-.090	.048	-.080	-.120	-.080	.068	-.090	-.110
	0	-.070	.074	-.134	-.114	-.144	.044	-.108	-.128	-.078	.054	-.054	-.086
	1	-.124	.086	-.170	-.108	-.074	.066	-.108	-.116	-.076	.068	-.080	-.082
	2	-.132	.076	-.150	-.114	-.080	.068	-.114	-.120	-.076	.066	-.066	-.092
	4	-.124	.076	-.118	-.108	-.110	.050	-.094	-.120	-.088	.076	-.098	-.106
3	-1	-.100	.086	-.126	-.108	-.088	.062	-.086	-.112	-.070	.082	-.074	-.098
	0	-.166	.076	-.162	-.112	-.118	.050	-.108	-.108	-.070	.068	-.068	-.084
	1	-.132	.092	-.166	-.110	-.138	.050	-.100	-.124	-.080	.070	-.070	-.096
	2	-.096	.078	-.110	-.118	-.096	.042	-.118	-.120	-.086	.050	-.090	-.104
	4	-.090	.074	-.128	-.110	-.112	.042	-.080	-.118	-.100	.066	-.066	-.098

Note. Skew = skewness of the level-1 error variance, Kurt = kurtosis of the level-1 error variance, ReML = restricted maximum likelihood estimation, MCMC = Markov chain Monte Carlo estimation, W = within-series model, B = between-series model, *I* = number of measurement occasions, *J* = number of participants.

Table 23.
RMSE of Level-2 Error Variance for the Treatment Effect in the Within-Series Model

		I=10				*I=20*				*I=40*			
		J = 4		*J = 8*		*J = 4*		*J = 8*		*J = 4*		*J = 8*	
Skew	Kurt	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC	ReML	MCMC
0	-1	.718	.134	.464	.126	.591	.179	.387	.176	.512	.217	.344	.228
	0	.707	.138	.469	.126	.605	.197	.391	.195	.509	.243	.345	.221
	1	.675	.126	.451	.138	.597	.184	.383	.184	.508	.219	.336	.226
	2	.704	.118	.452	.134	.591	.249	.382	.195	.504	.221	.332	.219
	4	.674	.126	.455	.130	.591	.176	.373	.184	.517	.228	.333	.224
1	-1	.666	.126	.453	.130	.574	.235	.394	.190	.505	.247	.335	.226
	0	.680	.134	.462	.126	.613	.207	.379	.192	.539	.266	.332	.221
	1	.700	.130	.454	.141	.566	.182	.381	.192	.520	.263	.333	.224
	2	.691	.145	.466	.141	.604	.184	.395	.184	.515	.241	.335	.219
	4	.670	.126	.458	.126	.607	.187	.404	.192	.503	.253	.332	.217
2	-1	.724	.118	.464	.118	.584	.167	.390	.182	.515	.228	.335	.205
	0	.714	.148	.460	.126	.602	.200	.383	.179	.500	.226	.344	.230
	1	.672	.192	.455	.155	.597	.226	.386	.190	.527	.245	.333	.224
	2	.688	.173	.454	.126	.602	.228	.387	.187	.519	.241	.342	.226
	4	.672	.130	.465	.126	.603	.195	.385	.184	.514	.255	.327	.219
3	-1	.705	.155	.463	.122	.609	.214	.391	.187	.519	.251	.327	.219
	0	.666	.155	.448	.126	.588	.192	.381	.202	.519	.237	.342	.230
	1	.663	.219	.437	.126	.588	.202	.375	.182	.532	.259	.332	.217
	2	.704	.141	.457	.126	.564	.184	.394	.187	.506	.217	.336	.217
	4	.709	.118	.459	.126	.579	.170	.394	.190	.525	.245	.332	.212

Note. Skew = skewness of the level-1 error variance, Kurt = kurtosis of the level-1 error variance, ReML = restricted maximum likelihood estimation, MCMC = Markov chain Monte Carlo estimation, W = within-series model, B = between-series model, *I* = number of measurement occasions, *J* = number of participants.

APPENDIX B. TABLES OF ETA-SQUARE ANALYSES

Table 24.
Eta-Square for Bias of the Treatment Effect Estimate

Within-Series Model		Between-Series Model	
Source	Eta-Square	Source	Eta-Square
Skewness*Kurtosis	.13	Skewness*Kurtosis	.07
I*Kurtosis	.03	I*Kurtosis	.06
J*Kurtosis	.03	J*Kurtosis	.04
I	.02	I*J	.03
Skewness	.02	I*Skewness	.03
I*Skewness	.02	J*Skewness	.02
Kurtosis	.02	Skewness	.02
I*J	.01	Kurtosis	.01
J*Skewness	.01	I	.00
Est	.00	J	.00
Kurtosis*Est	.00	Skewness*Est	.00
I*Est	.00	I*Est	.00
J	.00	Kurtosis*Est	.00
J*Est	.00	J*Est	.00
Skewness*Est	.00	Est	.00

Note. I = number of measurement occasions, J = number of participants, Est = estimation methods

Table 25.
Eta-Square for RMSE of the Treatment Effect Estimate

Within-Series Model		Between-Series Model	
Source	Eta-Square	Source	Eta-Square
I	.62	J	.93
J	.32	I	.07
I*J	.03	I*J	.00
Est	.02	Skewness*Kurtosis	.00
I*Est	.00	J*Kurtosis	.00
J*Est	.00	I*Kurtosis	.00
Kurtosis	.00	I*Skewness	.00
Skewness*Kurtosis	.00	Kurtosis	.00
I*Kurtosis	.00	J*Skewness	.00
I*Skewness	.00	Skewness	.00
J*Kurtosis	.00	Est	.00
J*Skewness	.00	Kurtosis*Est	.00
Skewness	.00	Skewness*Est	.00
Kurtosis*Est	.00	I*Est	.00
Skewness*Est	.00	J*Est	.00

Note. I = number of measurement occasions, J = number of participants, Est = estimation methods

Table 26.
Eta-Square for CI Coverage Rate of the Treatment Effect Estimate

Within-Series Model		Between-Series Model	
Source	Eta-Square	Source	Eta-Square
I	.41	I	.40
Est	.17	Est	.12
J*Est	.07	J	.04
Kurtosis	.02	I*Kurtosis	.04
I*Est	.02	Skewness*Kurtosis	.03
Skewness*Kurtosis	.01	I*Skewness	.02
I*Skewness	.01	I*J	.02
I*J	.01	Kurtosis	.02
I*Kurtosis	.01	J*Kurtosis	.02
J	.01	J*Est	.01
Kurtosis*Est	.00	J*Skewness	.01
Skewness	.00	I*Est	.00
J*Kurtosis	.00	Kurtosis*Est	.00
J*Skewness	.00	Skewness	.00
Skewness*Est	.00	Skewness*Est	.00

Note. I = number of measurement occasions, J = number of participants, Est = estimation methods

Table 27.
Eta-Square for CI Width of the Treatment Effect Estimate

Within-Series Model		Between-Series Model	
Source	Eta-Square	Source	Eta-Square
I	.57	J	.88
J	.40	I	.11
I*J	.02	I*J	.00
I*Est	.01	Est	.00
J*Est	.01	J*Est	.00
Est	.00	I*Est	.00
Skewness*Kurtosis	.00	Skewness*Kurtosis	.00
I*Skewness	.00	I*Kurtosis	.00
J*Kurtosis	.00	J*Skewness	.00
I*Kurtosis	.00	I*Skewness	.00
Kurtosis	.00	J*Kurtosis	.00
Kurtosis*Est	.00	Skewness	.00
Skewness*Est	.00	Kurtosis	.00
Skewness	.00	Kurtosis*Est	.00
J*Skewness	.00	Skewness*Est	.00

Note. I = number of measurement occasions, J = number of participants, Est = estimation methods

Table 28.
Eta-Square for Statistical Power of the Treatment Effect Estimate

Within-Series Model		Between-Series Model	
Source	Eta-Square	Source	Eta-Square
I	.55	J	.87
J	.41	I	.11
I*J	.03	I*J	.01
I*Est	.01	Est	.00
J*Est	.00	Skewness*Kurtosis	.00
Est	.00	I*Skewness	.00
Skewness*Kurtosis	.00	J*Kurtosis	.00
J*Kurtosis	.00	I*Kurtosis	.00
I*Skewness	.00	Skewness	.00
I*Kurtosis	.00	I*Est	.00
Skewness	.00	J*Est	.00
Kurtosis	.00	Kurtosis	.00
J*Skewness	.00	J*Skewness	.00
Kurtosis*Est	.00	Kurtosis*Est	.00
Skewness*Est	.00	Skewness*Est	.00

Note. I = number of measurement occasions, J = number of participants, Est = estimation methods

APPENDIX C. SAS CODES

Within-Series Model (ReML with Kenward-Roger)

```
proc mixed data=j1 covtest cl;
class idlevel2;
model y = phase time inter / s cl alpha = .05 ddfm = kenward;
random int phase time inter / sub = idlevel2;
repeated / sub = idlevel2;
run;
```

Within-Series Model (Bayesian)

```
proc mcmc data=j1  diag=all dic nbi=5000  nmc=20000 stats(alpha=(0.05  ))=(summary
intervals) dic monitor=(beta1 beta2 beta3 beta4 Sigma1 Sigma6 Sigma11 Sigma16 var_e);
ods output PostSummaries=esttmp3 PostIntervals=inttmp3 Geweke=geweke3
Heidelberger=heide3;

array Sigma [4,4];
array beta [4] ;
array b [4];
array mu0 [4] (0 0 0 0);
array Sig0[4,4] (1e6 0  0 0 0 1e6 0 0 0 0 1e6 0 0 0 0 1e6);
array SDIFFUSE[4,4] (1 0  0 0 0 1 0 0 0 0 1 0 0 0 0 1);
parms var_e   1;
parm  Sigma 1;
parms beta 0 ;
prior  beta   ~ mvn(mu0,Sig0);
prior var_e ~ igamma (shape=2.001, scale=1.001);
prior Sigma ~ iwish(6 , SDIFFUSE);
mu = (b1) + (b2)*phase + (b3)*time + (b4)*inter;
random b ~mvn(beta,Sigma) subject=idlevel2 ;
model y~normal(mu,var=var_e);
run;
```

Between-Series Model (ReML with Kenward-Roger)

```
* Homogeneous Variance Model;
proc mixed data =j1 covtest cl;
class idlevel2;
model y = D1 D1*phase P13 P23 P33 P1*phase / noint s cl alpha = .05 ddfm = kenward;
random D1 D1*phase / sub=idlevel2;
repeated / sub = idlevel2;
run;

* Hetrogeneous Variance Model;
proc mixed data =j1 covtest cl;
class idlevel2 Phcat1;
model y = D1 D1*phase P13 P23 P33 P1*phase / noint s cl alpha = .05 ddfm = kenward;
random D1 D1*phase / sub=idlevel2;
repeated / group = Phcat1 sub = idlevel2;
run;
```

Between-Series Model (Bayesian)

```
proc mcmc data=j1  diag=all dic nbi=5000  nmc=20000 stats(alpha=(0.05  ))=(summary
intervals ) dic monitor=(beta1 beta2 alpha1 alpha2 alpha3 alpha4 Sigma1 Sigma4 var_e);
ods output PostSummaries=esttmp4 PostIntervals=inttmp4 Geweke=geweke4
Heidelberger=heide4;

array Sigma [2,2];
array beta [2] ;
array b [2];
array alpha [4];
array mu0 [2] (0 0);
array mu1 [4] (0 0 0 0);
array Sig0[2,2] (1e6 0 0 1e6);
array Sig1[4,4] (1e6 0  0 0 0 1e6 0 0 0 0 1e6 0 0 0 0 1e6);
array SDIFFUSE[2,2] (1 0 0 1);
parms var_e   1;
parms Sigma 1;
parms beta 0 ;
parms alpha 0;
prior  beta  ~ mvn(mu0,Sig0);
prior  alpha ~ mvn(mu1,Sig1);
prior var_e ~ igamma (shape=2.001, scale=1.001);
prior Sigma ~ iwish(6 , SDIFFUSE);
mu = (b1)*D1 + (b2)*D1*phase + (alpha1)*P13 + (alpha2)*P23 + (alpha3)*P33 +
(alpha4)*P1*Phase;
random b ~ mvn(beta,Sigma) subject=idlevel2 ;
model y~normal(mu,var=var_e); run;
```