Statistical Properties of the Single Mediator Model with Latent Variables in the Bayesian Framework

by

Milica Miočević

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved February 2017 by the
Graduate Supervisory Committee:

Roy Levy, Co-Chair
David Mackinnon, Co-Chair
Kevin Grimm
Stephen West

ARIZONA STATE UNIVERSITY

May 2017

ABSTRACT

Statistical mediation analysis has been widely used in the social sciences in order to examine the indirect effects of an independent variable on a dependent variable. The statistical properties of the single mediator model with manifest and latent variables have been studied using simulation studies. However, the single mediator model with latent variables in the Bayesian framework with various accurate and inaccurate priors for structural and measurement model parameters has yet to be evaluated in a statistical simulation. This dissertation outlines the steps in the estimation of a single mediator model with latent variables as a Bayesian structural equation model (SEM). A Monte Carlo study is carried out in order to examine the statistical properties of point and interval summaries for the mediated effect in the Bayesian latent variable single mediator model with prior distributions with varying degrees of accuracy and informativeness. Bayesian methods with diffuse priors have equally good statistical properties as Maximum Likelihood (ML) and the distribution of the product. With accurate informative priors Bayesian methods can increase power up to 25% and decrease interval width up to 24%. With inaccurate informative priors the point summaries of the mediated effect are more biased than ML estimates, and the bias is higher if the inaccuracy occurs in priors for structural parameters than in priors for measurement model parameters. Findings from the Monte Carlo study are generalizable to Bayesian analyses with priors of the same distributional forms that have comparable amounts of (in)accuracy and informativeness to priors evaluated in the Monte Carlo study.

DEDICATION

For my parents. Thank you for raising me around math problems and eccentric people.

ACKNOWLEDGMENTS

I would like to express my most sincere appreciation to my committee chairs, Dr. David MacKinnon and Dr. Roy Levy, who were exceptionally supportive and dedicated mentors. Without your guidance and persistent help, this dissertation and this line of research would not have been possible. I would also like to thank my committee members, Dr. Steve West and Dr. Kevin Grimm, for their input during the development of this project. Your questions and comments were invaluable. I am grateful for Dr. Leona Aiken's guidance and advice; thank you for introducing me to the field, for having me as a teaching assistant for several courses, and for making the quant program a welcoming and supportive environment.

I am deeply thankful to my parents and my sister Olga for their positive outlook on life and good sense of humor. Thank you, also, to my aunt and uncle, Maja and Boba, for always being there when I need something. To my friends in the program: you are great colleagues and I feel lucky to have had you around, both in the office and during quant dinners. To my friends outside of the program: I truly appreciated our time together, particularly the trips, swims, races, bike rides, coffees, home-made meals, live shows, and hikes when things got stressful. And, of course, thank you to my dog Biggie for your company on the couch even when my work continued past your bedtime.

TABLE OF CONTENTS

Page

APPENDIX Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Sometimes a third variable can improve the understanding of the relation between two variables, X and Y. When a third variable is intermediate between X and Y in a causal chain, it is called a mediator (James & Brett, 1984; MacKinnon, 2008). Statistical mediation analysis consists of estimating the mediated effect, i.e., the indirect effect of X on Y through M. The single mediator model is the simplest model in statistical mediation analysis (Figure 1).

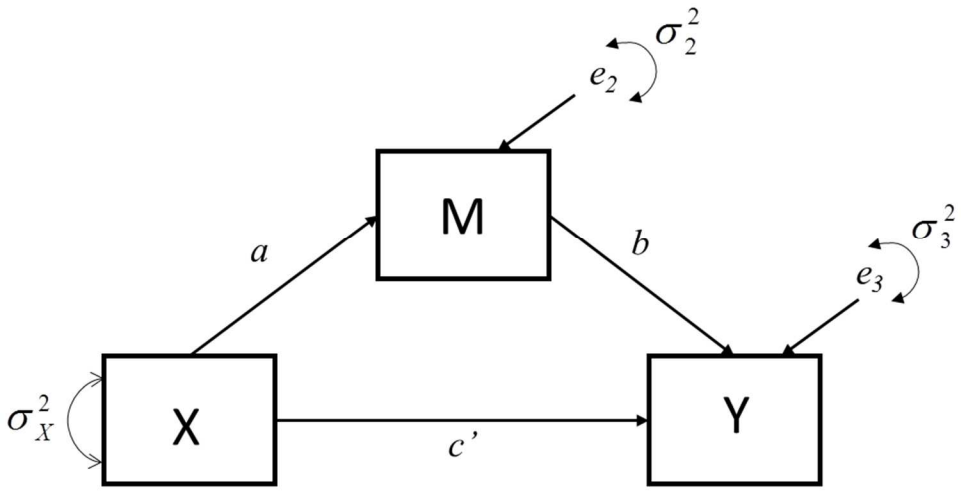

Figure 1. Single mediator model with manifest X, M, and Y.

With manifest variables, the single mediator model is often estimated using ordinary least squares (OLS) regression, and can be described using three equations (MacKinnon, 2008):

$$Y = i_1 + cX + e_1 \quad (1)$$

$$M = i_2 + aX + e_2 \quad (2)$$

$$Y = i_3 + c'X + bM + e_3 \quad (3)$$

where X is the independent variable, M is the mediator, and Y is the dependent variable. Intercepts are i_1, i_2, and i_3, *c* is the coefficient relating the independent variable to the dependent variable when the mediator is not taken into account, *a* is the coefficient relating the independent variable to the mediator, *b* is the coefficient relating the mediator to the dependent variable in the model containing the independent variable, *c'* is the coefficient relating the independent variable to the dependent variable in the model containing the mediator, and e_1, e_2, and e_3 are error terms assumed to follow a normal distribution with a mean of zero and variances of σ_1^2, σ_2^2, and σ_3^2 (respectively). In OLS regression with continuous variables the mediated effect is most often computed as the product of coefficients *ab*, although it can also be computed as the difference between coefficients *c*-*c'*. In models with random slopes the coefficients *a* and *b* could be correlated (MacKinnon, 2008; Tofighi, West, & MacKinnon, 2013), and the formula for the mediated effect becomes $ab + r_{ab}s_a s_b$, where r_{ab} is the correlation between coefficients *a* and *b*, s_a is the standard error of the *a* coefficient, s_b is the standard error of the *b* coefficient, and the product $r_{ab}s_a s_b$ equals the covariance of coefficients *a* and *b* (Craig, 1936). For uncorrelated coefficients *a* and *b*, i.e., $r_{ab} = 0$, the computation of the mediated effect simplifies to equal the product *ab* (Craig, 1936). The square root of the variance of the product of two correlated normal variates, here *a* and *b*, is as follows (Craig, 1936):

$$s_{ab} = \sqrt{a^2 s_b^2 + b^2 s_a^2 + 2abr_{ab}s_a s_b + s_a^2 s_b^2 + (r_{ab}s_a s_b)^2} \ . \qquad (4)$$

With uncorrelated *a* and *b* the formula in Equation 4 becomes $s_{ab} = \sqrt{a^2 s_b^2 + b^2 s_a^2 + s_a^2 s_b^2}$.

The correlation between the *a* and *b* paths does not exist in the population where there is only one true value of *a* and only one true value of *b*, as is the case in the Monte Carlo

study in this dissertation. For this reason, the mediated effect in this project will be computed as the product *ab*.

The standard error of the mediated effect is used in the construction of a confidence interval for the mediated effect, along with the estimate of the mediated effect *ab* and the critical value from the appropriate distribution, denoted $z_{1-\alpha/2}$ in Equation 5:

$$CI = [ab - z_{1-\alpha/2}s_{ab}, ab + z_{1-\alpha/2}s_{ab}]. \quad (5)$$

For normal theory confidence limits the same critical value $z_{1-\alpha/2}$ is used to compute both the lower and the upper confidence limit. However, the product of two normal distributions is not normal (Lomnicki, 1967; Springer & Thompson, 1966); instead it is symmetric with a kurtosis of six when the two variables have means equal to zero, and skewed with excess kurtosis when the two variables have means different from zero (Craig, 1936). For this reason, the confidence limits for the mediated effect based on the normal distribution have worse properties than confidence limits based on the distribution of the product and methods that make no distributional assumptions, such as the bootstrap (Cheung 2007, 2009; MacKinnon, Fritz, Williams, & Lockwood 2007; MacKinnon, Lockwood, & Williams, 2004; MacKinnon, Lockwood, Hoffmann, West, & Sheets, 2002; MacKinnon, et al., 1995; Shrout & Bolger, 2002; Tofighi & MacKinnon, 2011; Valente, Gonzalez, Miočević, & MacKinnon, 2015).

Unlike the normal distribution, the distribution of the product is not symmetric, and thus the critical values used to compute the upper and lower confidence limits often have different absolute values. The calculation of the confidence intervals for the mediated effect based on the distribution of the product has been simplified with programs called PRODCLIN and RMediation (MacKinnon, Fritz, Williams, &

Lockwood, 2007; Tofighi & MacKinnon, 2011). The bootstrap makes no distributional assumptions about the product *ab*, and can thus accommodate the nonnormality of this distribution and give asymmetrical confidence limits for the mediated effect (Manly, 1997; Shrout & Bolger, 2002; MacKinnon, Lockwood, & Williams, 2004; MacKinnon, 2008). However, the bias-corrected bootstrap confidence limits for the mediated effect have elevated Type I error rates when the non-zero path is medium or large, and power stagnates when the *a* path increases and the *b* path is small (Fritz, Taylor, & MacKinnon, 2012). The percentile bootstrap has comparable statistical properties to the distribution of the product method (MacKinnon, Lockwood, & Williams, 2004). Given the computational burden of bootstrap methods, the excessive Type I error rate of the bias-corrected bootstrap in some parameter combinations, and the comparable statistical properties of the percentile bootstrap and the distribution of the product method, the distribution of the product method will be the only frequentist interval estimation method in this Monte Carlo study.

Single Mediator Model with Latent Variables

With manifest (observed) X, M, and Y the single mediator model can be estimated in OLS regression. If at least one of the constructs in the model is latent and measured with multiple variables (indicators), then the simultaneous estimation of a measurement model for the latent variable and a structural model for latent variables requires using structural equation modeling (SEM). Modeling latent variables as manifest sums of their indicators can lead to the underestimation of the mediated effect due to measurement error (Hoyle & Kenny, 1999; MacKinnon, 2008; Ledgerwood & Shrout, 2011), thus it is recommended to estimate such models in the SEM framework.

SEM is a family of statistical methods for modeling and evaluating relations between variables. The main advantages of SEM over other methods stem from the possibility of modeling latent variables, the simultaneous estimation of the measurement model (the relations between indicators and latent variables) and structural model (the relations between latent variables), and the possibility of having one variable be both a predictor and an outcome (Hoyle, 2012). The single mediator model with latent variables is estimated in the SEM framework, and has already been studied using simulation studies. There are two articles on the topic of mediation analysis with latent variables in the frequentist SEM framework that are relevant for the Monte Carlo study in this dissertation (Finch, West, & MacKinnon, 1997; Falk & Biesanz, 2015). The following section will highlight the most relevant design choices and findings in these articles.

The single mediator model with latent variables and three indicators per latent variable, as described by Finch, West, and MacKinnon (1997), is the focus of this dissertation. Results of simulation studies using classical methods to fit this model have found that point estimates of the indirect and direct effects obtained using maximum likelihood (ML) and the asymptotically distribution free (ADF) method have less than 10% relative bias for sample sizes ranging from 150 to 1000, however, the standard errors were affected by the nonnormality of the indicators (Finch, West, & MacKinnon, 1997). The authors did not elaborate on how their findings inform confidence interval computation; however, by looking at the formula for the confidence interval of the mediated effect in Equation 5, it is clear that negatively biased standard errors can result in a confidence interval that is too narrow, which might affect coverage of the interval. The Monte Carlo study in this project will not evaluate effects of nonnormality of

indicators. However, Finch and coauthors selected model generating parameter values based on values commonly encountered in community research (Finch, West, & MacKinnon, 1997), so the same values are used for three out of four parameter combinations in this study. The fourth parameter combination was designed specifically for this study in order to investigate Type I error rates.

The second set of relevant findings for the design of the Mote Carlo study in this project come from a study that evaluated statistical properties of 11 estimators of the mediated effect in latent variable models (Falk & Biesanz, 2015). The 11 estimators were compared based on power, Type I error rates, and 95% coverage. Findings indicated that the best methods for both interval estimation and inference are the likelihood-based confidence intervals, the percentile bootstrap, and the distribution of the product method. The findings of Falk and Biesanz (2015) suggest that the percentile bootstrap and the distribution of the product are two frequentist methods that ought to be used to assess the relative performance of Bayesian credibility intervals.

Bayesian Mediation Models with Manifest Variables

Frequentist and Bayesian statistics differ in their definition of probability, which in turn leads to different approaches to statistical analysis. In frequentist statistics, probability is seen as the long run frequency of an event, parameters are treated as fixed while data are seen as random, and the interpretation of estimates rests on the assumption of repeated sampling which is almost never satisfied. In Bayesian statistics, probability is seen as a degree of belief, data are treated as fixed while parameters are assigned distributions, and the result of an analysis is a posterior distribution used to construct point and interval summaries with probabilistic interpretations. Bayesian statistical

analysis consist of applying Bayes' theorem to update the prior distribution of a set of parameters θ with observed data, modeled as $p(data|\theta)$, in order to obtain the posterior distribution, $p(\theta|data)$. Bayes' theorem simply states that the posterior distribution equals the product of the prior distribution and the likelihood function divided by the marginal distribution of the data: $p(\theta \mid data) = \frac{p(\theta)\,p(data \mid \theta)}{p(data)}$. Since $p(data)$ is a normalizing constant, this term can be omitted in order to obtain the following proportional relationship between the posterior and the product of the prior and the likelihood: $p(\theta \mid data) \propto p(\theta)\,p(data \mid \theta)$. The biggest controversy related to Bayesian statistics is the introduction of a prior distribution into the statistical analysis. However, if the prior distribution is diffuse (i.e., communicates no information), the numerical estimates in a Bayesian and frequentist analysis are often identical. The Bayesian single mediator model with latent variables builds on the Bayesian single mediator model with manifest variables. This section outlines steps in a Bayesian mediation analysis with manifest variables using two approaches (Yuan & MacKinnon, 2009; Enders, Fairchild, & MacKinnon, 2013). It is also possible to test for the presence of mediation using a third approach (Nuijten, Wetzels, Matzke, Dolan, & Wagenmakers, 2015), however, the focus of this dissertation is on parameter estimation and not hypothesis testing, so this approach is not described.

The goals of a Bayesian mediation analysis are the same as the goals of a classic mediation analysis: to determine whether a mediated effect is present, and/or to compute its numerical value. The mediated effect, denoted *ab* in the single mediator model with manifest variables, does not get assigned a prior distribution directly. Instead, priors are

assigned either to coefficients and error variances in equations 2 and 3 (Yuan & MacKinnon, 2009), or to the covariance matrix of variables X, M, and Y (Enders, Fairchild, and MacKinnon, 2013). In the following paragraphs, the two approaches are referred to as the method of coefficients (Yuan & MacKinnon, 2009), and the method of covariances (Enders, Fairchild, & MacKinnon, 2013).

In the method of coefficients the analysis starts by assigning prior distributions to regression coefficients, intercepts, and error variances (Yuan & Mackinnon, 2009). Common options for regression coefficients and intercepts are uniform and normal priors, however uniform priors state that all values outside of the minimum and maximum values of the range are impossible and thus normal priors are recommended instead in order to incorporate more uncertainty into the analysis. Some possible priors for the variance parameter are the inverse-gamma prior, or a uniform prior if limiting values for the variance are available from prior sources (Yuan & MacKinnon, 2009). Another option would be to specify a joint uniform prior ranging between $-\infty$ and ∞ for parameters *a*, *b*, and the log of the variance parameter. This prior is improper, but usually results in a proper posterior and correct inferences (Yuan & MacKinnon, 2009). The results of a Monte Carlo study indicated that both frequentist and Bayesian methods for the single mediator model are unbiased, as shown by values of empirical bias and relative mean square error (MSE) (note that this is not the same as the Root Mean Square Error, a criterion that is used in the Monte Carlo study for this project). Without prior information, Bayesian methods have relative MSE values comparable to those of frequentist methods, and with prior information the relative MSE values of Bayesian methods are noticeably smaller those of frequentist methods, especially at sample sizes

smaller than 200. Coverage of credible intervals tended to be higher than the nominal value when sample sizes were 200 and 1000 and the mediated effect was small, when the mediated effect was large and the sample size was 25 or 50, and when the mediated effect was zero (Yuan & MacKinnon, 2009).

In the method of covariances the analysis starts by assigning a prior distribution to the covariance matrix of X, M, and Y (Enders, Fairchild, & MacKinnon 2013). The authors proposed an inverse-Wishart prior for the covariance matrix, and computed values of coefficients and residual variances in equations 2 and 3 using elements from the posterior covariance matrix:

$$a = \frac{\sigma_{XM}}{\sigma_X^2} \tag{6}$$

$$b = \frac{\sigma_{MY}\sigma_X^2 - \sigma_{XM}\sigma_{XY}}{\sigma_X^2\sigma_M^2 - \sigma_{XM}^2} \tag{7}$$

$$c' = \frac{\sigma_{MY}\sigma_X^2 - \sigma_{XM}\sigma_{XY}}{\sigma_X^2\sigma_M^2 - \sigma_{XM}^2} \tag{8}$$

$$\sigma_{\varepsilon_M}^2 = \sigma_M^2 - \frac{\sigma_{XM}^2}{\sigma_X^2} \tag{9}$$

$$\sigma_{\varepsilon_Y}^2 = \frac{\sigma_X^2\sigma_M^2\sigma_Y^2 - \sigma_X^2\sigma_{MY} - \sigma_M^2\sigma_{XM}^2 - \sigma_Y^2\sigma_{XM}^2 + 2\sigma_{XM}\sigma_{XY}\sigma_{MY}}{\sigma_X^2\sigma_M^2 - \sigma_{XM}^2} \tag{10}$$

Moving from the manifest to the latent variable framework requires the addition of a measurement model for all latent variables. The single mediator model with latent variables can also be estimated using the method of covariances and the method of coefficients, however, this project only uses the method of coefficients because it is more

intuitive for substantive researchers to think about their prior knowledge in terms of structural paths and loadings than covariances between latent variables and manifest indicators. As of November 2016, there is only one published paper describing and evaluating the statistical properties of the mediated effect in a model similar to the one in this dissertation (Chen, Choi, Weiss, & Stapleton, 2014). Chen and colleagues (2014) examined the bias of point summaries and coverage of the equal-tail credibility intervals of the mediated effect in the single mediator model with latent variables with complete mediation at N = 50, 100, and 400. The findings indicate that for smaller sample sizes the maximum likelihood point estimates are preferred in terms of bias and the bias-corrected bootstrap has coverage closer to nominal levels of 0.95 than do MCMC methods, however, at N = 400 and with larger effects MCMC methods tend to produce intervals with coverage closer to the nominal level of 0.95 than the coverage of the bias-corrected bootstrap confidence intervals. Note that Chen and colleagues (2014) evaluated only diffuse prior distributions, and that not all priors were conjugate, i.e., uniform priors ranging from -1 to 1 were used for structural paths and loadings. Thus, even the condition with diffuse priors in the Monte Carlo study for this project adds new information about statistical properties of Bayesian methods, since this project makes use of conditionally conjugate (i.e., normal) priors for loadings and structural paths.

The next section gives a general introduction to Bayesian SEM; this section is included for completeness, however, it is not required to understand the Monte Carlo study in this project. The subsequent section will describe the single mediator model with latent variables conceptualized as a Bayesian SEM; the content of this section is directly relevant to the Monte Carlo study in this project.

Bayesian Structural Equation Models (SEMs)

SEMs are commonly estimated using ML estimation, and interpreted using frequentist terms and definition of probability. It is also possible, and sometimes more advantageous (Lee & Song, 2004; Rindskopf, 2012) to fit SEMs in a Bayesian framework (Kaplan & Depaoli, 2012; Levy & Choi, 2013; Scheines, Hoijtink, & Boomsma, 1999). The following sections briefly outline how Bayesian SEMs are specified, estimated, and evaluated. For more general and extensive accounts of Bayesian SEM, see chapters by Kaplan and Depaoli (2012), and Levy and Choi (2013).

Model Specification

By specifying a model the researcher is formalizing the mechanisms hypothesized to have produced the observed data. SEMs have a measurement model and a structural model. Building on methods that novices to SEM might be comfortable with, Song and Lee (2012) describe the measurement model as a confirmatory factor analysis, and the structural model as a regression equation with latent variables. In Bayesian SEM all parameters in the structural and measurement models have distributions. Before the analysis, parameters are assigned a prior distribution, and after the prior has been updated with the observed data, inferences about parameters are drawn from the posterior distribution. The prior distributions can have various levels of informativeness that exist on a spectrum, but are often labeled as either noninformative (also referred to as vague or diffuse) or informative. It is also possible to specify so-called weakly informative priors, which contain more information than diffuse priors, but do not reflect the actual amount of prior knowledge/ intuition the researcher possesses (Gelman, Carlin, Stern, & Rubin, 2004). There is more than one way of conceptualizing what weakly informative means,

namely, a prior can be weakly informative if it is more dispersed (has a larger spread) than a prior elicited from expert opinion, or as defined by Evans and Jang (2011) a weakly informative prior is one that has less prior-data conflict (which occurs when the bulk of the prior density is in a space where the likelihood function is low) than an elicited prior that reflects the current state of knowledge. Another possible non-informative prior specification is the so-called *unit information prior*, which carries the amount of prior information equivalent to what can be obtained from a prior sample size of 1 (Kass & Wasserman, 1995). With noninformative priors the point summaries (mean, median, and mode) of posterior distributions of the parameters are often numerically close to estimates obtained using ML. Another important consideration in the choice of prior is conjugacy: if a conjugate prior is selected, then the prior and the posterior distributions are from the same family. Conjugacy is desirable because it makes it easier to obtain analytical solutions, which was important historically, but is no longer necessary for model estimation since the invention of Markov Chain Monte Carlo (MCMC) methods for approximating the posterior. Both informativeness and conjugacy of prior distributions will be described in more detail in the section Bayesian Mediation with Latent Variables, as they are both crucial for the design of prior distributions for the Bayesian methods in the Monte Carlo study in this dissertation.

The normal distribution is a conditionally conjugate prior distribution for measurement and structural intercepts, factor loadings, and structural coefficients in SEM, meaning that conditional on other model parameters this prior specification leads to posteriors that are also normal distributions (Kaplan & Depaoli, 2012). Covariance matrices of uniquenesses and structural disturbances are usually assigned inverse-gamma

distributions, if modeled individually, or an inverse-Wishart distribution, if modeled jointly. The inverse-gamma (IG) is a conditionally conjugate prior for variance parameters, meaning that if an inverse gamma prior is assigned to the variance parameter of a variable that is modeled as following a normal distribution conditional on the variance parameter, then the posterior for the variance, conditional on other model parameters, is also an inverse-gamma (Gelman, 2006).

The parameters of a prior distribution are called hyperparameters. The hyperparameters of a normal prior distribution are the mean (location) parameter, and the scale parameter (usually the variance, precision, or standard deviation, depending on the software). The inverse-gamma (IG) distribution is specified using the shape (denoted α), and scale (denoted β) parameters. Since the hyperparameters of an inverse-gamma prior are not as easy to conceptualize as the hyperparameters of a normal prior, one can specify α and β by imagining a prior sample was observed. Using this approach, α equals one half of the variance of the prior pseudo-sample, and β equals one half of the product of the variance of the prior pseudo-sample and the sample size of the pseudo-sample (Gelman, Carlin, Stern, & Rubin, 2004). Thus, an inverse-gamma prior for a parameter can be specified either as:

$$\theta \sim IG(\alpha, \beta) \tag{11}$$

where θ is the parameter being assigned a prior, α is the shape parameter, and β is the scale parameter, or as

$$\theta \sim IG(\nu_0 / 2, \nu_0 \sigma_0^2 / 2) \tag{12}$$

where ν_0 is the sample size of the pseudo-prior sample, and σ_0^2 is the variance of the prior pseudo-sample. A frequent recommendation for specifying a diffuse prior for the variance is to set the shape (α) and scale (β) parameters of the inverse-gamma to 0.001. However, the mean and variance for the inverse gamma with a scale parameter smaller than 1 and 2 (respectively) cannot be computed (Gelman, Carlin, Stern, & Rubin, 2004), and an *IG*(0.001, 0.001) prior can be problematic for variance parameters in hierarchical models (Gelman, 2006).

The most common way to model covariance matrices (of uniquenesses, disturbances, and sometimes variables) is by using an inverse-Wishart prior. The hyperparameters of an inverse-Wishart are a degrees of freedom parameter, ν , and a symmetric, positive definite matrix S.

$$p(\Sigma) \sim W^{-1}(\nu, S) \tag{13}$$

In Equation 13 the degrees of freedom parameter ν is often thought of as the prior sample size, and the S matrix is often set equal to the matrix of sums of squares and cross-products. The inverse-Wishart with a low degrees of freedom parameter (at least equal to the number of variables) and a fixed scale matrix S is often used as a reference (non-informative) proper prior for the covariance matrix (Daniels & Kass, 1999). Even though the inverse-Wishart distribution is a conjugate prior for a covariance matrix, it is also restrictive in that it only allows for one parameter that quantifies the degrees of freedom, which implies the same amount of prior information for all elements of the covariance matrix (Gelman, Carlin, Stern, & Rubin, 2004). There have been several proposed reparameterizations of the covariance matrix along with priors for the parameters

(Barnard, McCulloch, & Meng, 2000; Daniels & Kass, 1999, 2001; Leonard & Hsu, 1992; Yang & Berger, 1994), however, those will not be discussed here.

Model Estimation

Markov Chain Monte Carlo (MCMC) methods are necessary to estimate parameters in Bayesian SEMs because closed-form solutions are generally not available (Levy & Choi, 2013). Both ML and MCMC are iterative procedures, however, while ML seeks to maximize the likelihood function, MCMC seeks to approximate the entire joint posterior distribution of the parameters. MCMC methods are useful when it is not possible to sample parameter values directly from the posterior distribution, and a method that can approximate the posterior is needed. The target distribution of MCMC is set to be the posterior distribution of interest, and once the chains converge to the stationary (target) distribution, MCMC operates by using conditional distributions to draw values that, in limit, may be taken as draws from the desired distribution. These draws ultimately result in a joint (posterior) distribution of all model parameters. The key to the success of MCMC is not the Markov property but the improvement of the approximate distribution at each step of the simulation in the sense of convergence to the target distribution (Gelman, Carlin, Stern, & Rubin, 2004).

If the conditional distribution is of a known form, then Gibbs sampling (also called alternating conditional sampling) can be used to approximate the posterior distributions (Gelman, Carlin, Stern, & Rubin, 2004). The Gibbs sampler is a computer-intensive method for taking draws from a marginal distribution without having to calculate its density (Casella & George, 1992). The Gibbs sampler uses the full set of univariate conditionals to define a single iteration, meaning that all parameters for which

a conditional value is not being calculated are being conditioned on. Casella and George (1992) describe the Gibbs sampler as a practical application of the fact that conditional distributions are sufficient to determine the joint distribution, if it exists.

A chain that moves well through the space of the posterior distribution is said to mix well. Initial draws from the posterior are referred to as burn-in iterations, and are discarded before the posterior distribution is summarized. One issue with iterative methods, MCMC included, is the within-chain correlation of the draws. Correlated draws make assessing convergence using the Potential Scale Reduction factor (PSR, described in the next paragraph) less precise (Lee, 2007). Autocorrelation between draws is dealt with by thinning, and by having multiple chains run in parallel (Levy & Choi, 2013). Thinning is the practice of retaining only every i^{th} draw, where the thinning parameter i is chosen so that the retained draws are independent. However, thinning is not necessary, "as long as a sequence has converged and the number of iterations retained is substantial, it makes no practical difference if we keep all or every 25^{th} or every 50^{th} iteration" (Scheines, Hoijtink, & Boomsma, 1999). Furthermore, in some situations not thinning leads to more accurate values of the posterior mean and smaller standard deviations of the posterior (Link & Eaton, 2012).

There is no way of confirming that a chain has converged to the target (here, posterior) distribution, but there are several techniques one can use to find evidence of convergence. Cowles and Carlin (1996) reviewed all of the available convergence diagnostics at the time, and more recently Sinharay (2004) conducted a similar review of convergence diagnostics for psychometrics. In the social sciences literature, the most commonly encountered convergence diagnostics are those offered by the majority of

software packages, which are the Potential Scale Reduction factor (Gelman & Rubin, 1992), Geweke's diagnostic (1992), and trace plots of draws plotted against the iteration number for each parameter (Brooks, 1998).

Potential Scale Reduction (PSR) factor is computed as the square root of within and between chain variance divided by within chain variance (Gelman & Rubin, 1992; Brooks & Roberts, 1998).

$$PSR = \sqrt{\frac{W + B}{W}} \tag{14}$$

Values of PSR slightly above 1 (preferably below 1.1 according to Gelman et al., 2004) are considered evidence of convergence. In other words, convergence is achieved when there is little between-chain variance relative to within-chain variance. One criticism of PSR is that choosing overdispersed starting values (relative to the target distribution) for the chains requires knowledge about the target distribution, which the user does not have (Cowles & Carlin, 1996). Furthermore, Muthén and Asparouhov (2012) caution that PSR convergence after a certain number of iterations can be negated with additional iterations, and recommend that chains be run for longer after PSR convergence in order to assure that PSR remains only slightly larger than 1 even after additional iterations. Another convergence diagnostic measure was developed by Geweke (1992) and is calculated by computing the difference of the means of the first n_A iterations and the last n_B iterations ($n_A+n_B < n$, where n is equal to the total number of iterations) and dividing it by the square root of the asymptotic variance of the chain. As Cowles and Carlin (1996) point out, "by the central limit theorem, the distribution of this diagnostic approaches a

standard normal as n tends to infinity". Geweke (1992) recommends using $n_A = .1n$ and $n_B = .5n$ in the computation, but makes no recommendations about potential cut-off values that would indicate convergence. Zhang (2013) interpreted absolute values of the Geweke statistic below 1.96 as evidence of convergence. Another problem with Geweke's diagnostic measure is its sensitivity to the choices of n_A and n_B. One point made by Cowles and Carlin (1996) that has not changed to this day is that "automated convergence monitoring (as by a machine) is unsafe and should be avoided."

Other issues to consider with MCMC are the number of chains, starting values, and determining burn-in and stopping time (Gilks, Richardson, & Spiegelhalter, 1996). If the chain is irreducible, meaning that from any point in the distribution the chain has a positive probability of reaching any other point, the choice of starting values for a single chain does not matter very much (Roberts, 1996). It is a good idea to run at least three chains (Gelman & Shirley, 2011), and to have dispersed starting values for different chains in order to assist with the monitoring of convergence (Gelman & Rubin, 1992). Stopping time refers to ending the sampling and depends on time constraints, how long the chain(s) ran before convergence, the researcher's confidence that convergence was reached, and the autocorrelation between draws (and thus how much thinning one chooses to do, and how many iterations one wishes to use for inference). Comparing the number of iterations to ESS, effective sample size (defined as the sample size if there was no autocorrelation), is a proxy measure for autocorrelation; the closer the ESS is to the number of iterations, the lower the autocorrelation in the chain (Zhang, 2013). A rule of thumb is to have ESS of at least 400. Once approximate convergence has been reached, Gelman and Shirley (2011) recommend mixing all of the non-discarded draws from all

chains (the authors suggest discarding the first half of each chain and using the second halves for this procedure) in order to summarize the target (posterior) distribution. In this project three chains will be run for each Bayesian analysis, and preliminary analyses of single samples for each Monte Carlo study condition will be run in order to select the number of burn-in samples. Trace plots and the PSR will be used as criteria for diagnosing convergence.

MCMC is used for many Bayesian models with no analytical solutions, and thus the guidelines above apply regardless of whether one is seeking a posterior for a parameter as simple as a mean difference, or as complex as a cross-loading in SEM. There are several software options for specifying Bayesian SEMs, from programs for general Bayesian analysis that allow for the specification of SEMs such as WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000) also available online under the name WebBUGS (Zhang, 2014) and through a SAS interface (Zhang, McArdle, Wang, & Hamagami, 2008), various R packages (R core team, 2014), and the SAS procedure MCMC (SAS Institute Inc., 2009), to general SEM software that has Bayesian capabilities, such as AMOS (Arbuckle, 2007), Mplus (Muthén, & Muthén, 1998-2015), and Stata 14 (StataCorp, 2015). Stan (Stan development team, 2015), another program for Bayesian analysis, uses Hamiltonian Monte Carlo (HMC) to draw samples from the posterior (Neal, 2011; Stan development team, 2015). HMC chains tend to have low autocorrelation and to converge rapidly, and due to how the step in the chain is proposed, HMC draws tend to have high probability of acceptance which avoids the "exploration of the state space that occurs when Metropolis updates are done using a random-walk proposal distribution" (Neal, 2011). For more on HMC, see the Stan manual (Stan

development team, 2015). The R package R2WinBUGS (Sturtz, Ligges, & Gelman, 2005) allows for specifying the kinds of priors proposed for the Monte Carlo study in this project, and will thus be used for all Bayesian methods in this project.

Model Evaluation

As in frequentist estimation, model estimation is followed up with model evaluation and model comparison in situations where there are several competing models. According to Kaplan and Depaoli (2012) Bayesian SEMs can be evaluated using posterior predictive checking (PPC), and compared using Bayes Factors (BF), the Bayesian Information Criterion (BIC), and the Deviance Information Criterion (DIC). Model evaluation is not a part of the proposed Monte Carlo study, thus for the sake of brevity the model evaluation and comparison indices will not be covered in this document. For information on posterior predictive checking (PPC), see Meng (1994), Gelman and Meng (1996), and Gelman, Meng, and Stern (1996). Details about Bayes Factors can be found in Raftery (1993) and Kass and Raftery (1995). For situations where the computation of the BF is too complicated Kass and Raftery (1995) recommend approximating it by computing the Bayes Information criterion (BIC; Schwartz, 1978). The Deviance Information Criterion (DIC) is another useful index for comparing competing models (Spiegelhalter, Best, Carlin, & van der Linde, 2002), however with missing data there are several ways of computing the DIC depending on the chosen representation of the missing data structure (Celeux, Forbes, Robert, & Titterington, 2006; de la Torre & Douglas, 2008). Two additional indices for comparing models have been proposed, but not used as extensively as the measures above: the L_v (called L-

measure) for both parametric and semiparametric structural equation models (Song, Xia, Pan, & Lee, 2011), and model averaging described by Raftery (1993).

Bayesian Mediation with Latent Variables

In both frequentist and Bayesian frameworks with manifest variables it is possible to estimate single mediator models using regression and SEM. However, the single mediator model with latent variables needs to be specified as a SEM in order for the measurement and structural portions of the model to be estimated simultaneously.

Two ways of doing Bayesian mediation with the goal of estimating the mediated effect (as opposed to significance testing alone, which in the Bayesian framework can be done using Bayes Factors) involve specifying priors for regression coefficients (Yuan & MacKinnon, 2009), and specifying priors for the covariance matrix of variables X, M, and Y (Enders, Fairchild, and MacKinnon, 2013). One benefit of specifying prior distributions for the covariance matrix versus specifying priors for coefficients and residual variances is the ease of adding more mediators, moderators, and covariates to the model by simply increasing the dimensionality of the prior for the covariance matrix. Another benefit of the method of covariances is the ability to interpret the degrees of freedom parameter as a prior sample size, and thus control the weight the prior carried in the analysis through changes in the degrees of freedom parameter. However, the method of coefficients allows for a more intuitive way of thinking about the priors specified for the parameters, e.g., it is much easier to think of the expected value of the *a* path in the single mediator model than about the expected covariance between variables X and M and the expected variance of X needed to compute the *a* path (Equation 6). The

remainder of this section will focus on steps and considerations when fitting a single mediator model with latent variables as a Bayesian SEM.

In order to present the steps in estimation of a single mediator model with latent variables as a Bayesian SEM, a single data set was simulated inspired by a prevention study designed to evaluate and improve the health of law enforcement officers (Kuehl et al., 2014). The latent variables ξ, η_1, and η_2 stand for health, vitality, and activity (respectively). Each latent variable has three manifest indicators. The manifest indicators are scales (as opposed to scale items) with reliability 0.7. The model-generating parameter values equal those in combination 1 in the Monte Carlo study, and are as follows: all unstandardized loadings equal 1, all measurement error variances follow a normal distribution with a mean of 0 and a variance of 0.4286 (i.e., precision of 2.3332), and all three latent variables are normally distributed with means of 0 and variances (precisions) of 1. The structural paths equal $\gamma_{11} = 0.60$, $\beta_{21} = 0.20$, and $\gamma_{21} = 0.12$, and residual terms ζ_1 and ζ_2 follow a normal distribution with a mean of 0 and variances of 0.64 and 0.9168 (respectively) so latent variables η_1 and η_2 have variances (precisions) equal to 1. Note that for simulating data it was more intuitive to consider variances than precisions, however, R2WinBUGS uses the variance parametrization, thus the model fitting steps are discussed in terms of precision and not variance.

The first step of a Bayesian SEM analysis is the specification of prior distributions for all freely estimated model parameters. As in the frequentist framework, the scale of the latent variables needs to be set, and the options are either fixing one loading per latent variable to 1, or fixing the variances of latent variables to 1. In the Bayesian framework, setting one loading to 1 is the more common choice (Kaplan &

Depaoli, 2012). Setting the variance of a latent variable to 1 imposes the restriction that the sum of squared loadings and residual variances of manifest indicators must equal the variance of the manifest indicator (MacCallum, Edwards, & Cai, 2012), which is not always easy to consider when specifying priors for loadings and residual variances. This project makes use of conjugate priors for all model parameters, which are normal priors for loadings and structural paths, and gamma priors for precision parameters (Gelman, Carlin, Stern, & Rubin, 2004). The hyperparameters of the priors in this analysis were selected to communicate very little information. For loadings, it was assumed that all scales are equally reliable, thus the mean hyperparameter was set to equal the same value that the first loading was fixed at, i.e., 1. It was assumed that positive and negative value of structural parameters were equally likely, so the mean hyperparameter was chosen to be zero. It was assumed that the precision of the exogenous variable, the measurement error precisions, and the residual precisions of the endogenous variables all equal 1. However, the assumption that all precision parameters are equal to each other and all equal to 1 is not very plausible, so these priors were unit-informative (Kass & Wasserman, 1995), i.e., they were assigned a weight of 1 prior observation. For gamma prior distributions the shape (α) and scale (β) hyperparameters were set to $\alpha = ½ = \beta = (1)½$, thus representing the assumption that the best guess for the precision is 1, and that this value is given the weight of one observation (Gelman, Carlin, Stern, & Rubin, 2004). Two different approaches were used for determining the spread of normal priors. In the first scenario, the normal priors were assigned precision hyperparameters equal to .001, which is a very large spread given the expected values of loadings and structural paths in SEM. The priors in this scenario will be referred to as “diffuse generic priors” in the

remainder of the document. In the second scenario, the spread of normal priors was conditional on the corresponding precision parameter (for loadings, it was the measurement error precision of the manifest indicator, and for structural paths it was the residual precision of the endogenous variable). This specification yields a fully conjugate prior for the loading and corresponding measurement error precision, and the structural path and corresponding residual precision. These priors are referred to as "diffuse conjugate priors" in the remainder of the document.

The second step of a Bayesian analysis is to update the prior with the observed data. Since latent variable models do not have analytical solutions (Aitkin & Aitkin, 2005), this was done by using MCMC methods to approximate the posterior distribution. The use of MCMC requires specifying the number of chains (here 3 chains), burn-in iterations (1500 for these analyses), posterior draws after burn-in (3500 per chain, leading to 10500 total), and the thinning parameter (set to 1, i.e., no thinning for these analyses). The user can also specify starting values, as was done in this analysis, however, another option is to have WinBUGS generate starting values. In the case of mediation analysis, the parameter of interest is the mediated effect. In order to compute the marginal posterior of the mediated effect, the product $\gamma_{11}\beta_{21}$ was computed at each iteration, thus yielding an approximation of the marginal posterior based on 10500 draws from the three combined chains. The convergence of the chains was diagnosed using trace plots and PSR. Inferences can be made based on the marginal posteriors of the parameter(s) of interest, which in this case is the mediated effect.

The posterior for the mediated effect is summarized using the mean and median as point summaries, and equal-tail and highest posterior density (HPD) intervals as

interval summaries. With diffuse generic priors the posterior mean and median were equal to 0.13, and the 95% equal-tail (0.043, 0.233) and HPD (0.038, 0.227) intervals did not include zero. The analysis with diffuse conjugate priors yielded comparable point and interval summaries; the mean and median of the posterior were equal to 0.13, and the 95% equal tail (0.045, 0.232) and HPD (0.042, 0.229) intervals did not contain zero. Thus, the conclusion from analyses with both diffuse priors is that the mediated effect lies between 0.04 and 0.23 with 95% probability, and that the average of the posterior for the mediated effect equals .13. On average, one unit increase in health led to .13 units of increase in activity through vitality. Latent variables health, vitality, and activity are in units of the first indicator for each latent variable.

The Present Study

As of April 2016, there has been some published work suggesting the use of Bayesian methods to fit the single mediator model with latent variables in the SEM framework (Levy & Choi, 2013; Chen, Choi, Weiss, & Stapleton, 2014) as well as some literature on statistical properties of the single mediator model with latent variables estimated as a Bayesian SEM with continuous indicators (Chen, Choi, Weiss, & Stapleton, 2014) and with ordinal indicators (Chen, Zhang, & Choi, 2015). However, there are still several unanswered methodological questions about the statistical properties of point and interval summaries of the mediated effect in this model. First, it is unknown whether there are benefits of using diffuse normal priors in a Bayesian analysis relative to ML estimation (other than the probabilistic interpretation of the results). Second, the amount of improvement in Bayesian estimation of the mediated effect with accurate informative priors is unknown. Third, the detrimental effects of inaccurate

informative priors on Bayesian estimation of the mediated effect have yet to be quantified. Fourth, it is not known how the detrimental effects of inaccurate priors differ for measurement and structural model parameters. This dissertation examined these questions in a Monte Carlo study.

Before describing the Monte Carlo study, it is important to define how the terms *weight* and *direction of bias* are used in this project. The prior and the likelihood function both contribute certain amount of information to the posterior. Here, *weight* refers to the relative contribution of the data and the prior to the posterior; e.g., saying that the prior has 25% of the weight of the likelihood means that the prior is contributing 4 times less information than the likelihood. *Direction* of bias is used synonymously with *sign* of bias; e.g., saying that the *direction* of bias in the posterior is the same as the *direction* of the bias in the prior means that both the prior and the posterior had either positive bias (the expectation about a parameter was higher than the true value) or negative bias (the expectation about a parameter was lower than the true value).

CHAPTER 2

MONTE CARLO STUDY

This Monte Carlo study explores four related questions pertaining to the statistical properties of the mediated effect in the single mediator model with latent variables: 1) are there benefits or problems with using Bayesian methods with diffuse priors instead of ML, 2) how pronounced are the benefits of using Bayesian methods in the best case scenario where informative priors for measurement and structural parameters are accurate and carry 50% of the weight of the likelihood, 3) how large are the risks of using Bayesian methods in the worst case scenario where informative priors for measurement and structural parameters are inaccurate and carry 50% of the weight of the likelihood, and 4) is the change in statistical properties with the use of inaccurate priors more problematic for structural model parameters or measurement model parameters? Research questions 1 and 4 compare two options, and naturally lend themselves to being phrased as testable hypotheses. Research questions 2 and 3 deal with the extent to which statistical properties change with accuracy and informativeness in prior distributions, and are not appropriately answered by generating hypotheses. In the text below, questions 1 and 4 have been converted to testable hypotheses, and questions 2 and 3 remained research questions.

Hypotheses and Research Questions

The first hypothesis deals with a comparison between Bayesian methods with diffuse priors to ML point estimates and distribution of the product interval estimates of the mediated effect. It is hypothesized that point and interval summaries from Bayesian

methods with diffuse priors will have comparable statistical properties to ML point estimates and interval estimates using the distribution of the product.

The second research question focuses on the improvement of statistical properties in the "best case scenario" in this study, i.e., when Bayesian methods are used with accurate informative priors, relative to statistical properties using ML and distribution of the product. Bayesian methods with most informative accurate priors for structural and measurement model parameters are expected to have the best statistical properties out of all methods in the study. However, the amount of improvement in bias and efficiency for point summaries, and power, Type I error rate, coverage, imbalance, and interval width for interval summaries is unclear. This study will explore the amount of improvement in statistical properties of point and interval summaries of the mediated effect with the addition of accurate prior information that carries 25% and 50% of the weight of the likelihood function.

The third research question examines the "worst case scenario" in this study, i.e., the statistical properties of the mediated effect when Bayesian methods are used with inaccurate informative priors. Bayesian methods with the most informative inaccurate priors for the structural and measurement model parameters are expected to have the worst statistical properties out of all methods in the study. However, the amount of decline in statistical properties of the mediated effect with inaccurate priors is unknown. This study will quantify the risks of using inaccurate informative priors that have the 25% and 50% of the weight of the likelihood function.

The fourth hypothesis deals with the impact of inaccuracy in priors for structural paths and loadings. The mediated effect is computed as the product of two structural

paths. Thus, it is expected that the prior information for structural paths will have more bearing on the statistical properties of the mediated effect than prior information for measurement model parameters. More specifically, it is hypothesized that the (in)accuracy and informativeness of priors for structural paths will have more bearing on the statistical properties of the point and interval summaries of the mediated effect than the (in)accuracy and informativeness of priors for loadings.

Method

Data

The single mediator model with latent variables and three indicators per latent variable, depicted in Figure 2, is the model studied in this research. The mediated effect is computed as the product of structural paths $\gamma_{11}\beta_{21}$.

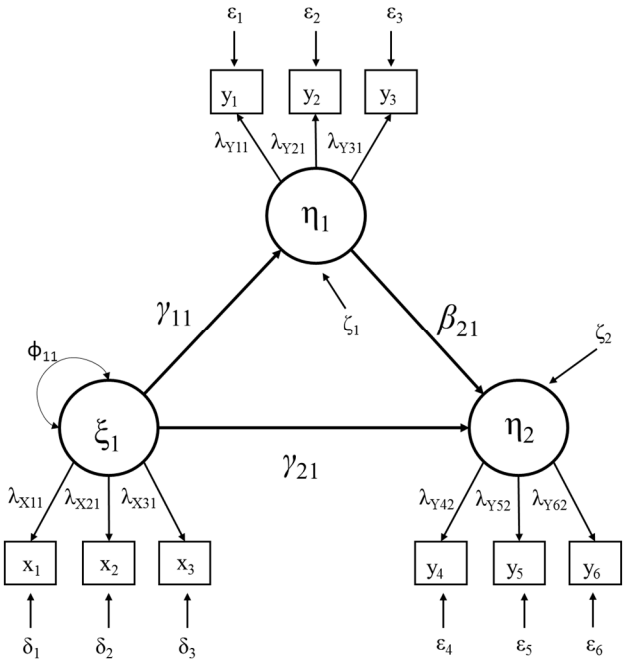

Figure 2. Single mediator model with latent X, M, and Y, and three manifest indicators per latent variable.

The model is described using equations 15-17:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \lambda_{X11} \\ \lambda_{X21} \\ \lambda_{X31} \end{bmatrix} \xi_1 + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix} \tag{15}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} = \begin{bmatrix} \lambda_{Y11} & 0 \\ \lambda_{Y21} & 0 \\ \lambda_{Y31} & 0 \\ 0 & \lambda_{Y42} \\ 0 & \lambda_{Y52} \\ 0 & \lambda_{Y62} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix} \tag{16}$$

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \beta_{21} & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} \\ \gamma_{21} \end{bmatrix} \xi_1 + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} \tag{17}$$

Data were simulated in R (R Core Team, 2014) according to equations 15-16 for the measurement model, and equation 17 for the structural model. Appendix A contains analytical formulas for the covariance matrix of manifest and latent variables for the model. Measurement errors of manifest indicators were simulated to follow normal distributions with a mean of 0 and variances equal to .4286. The disturbances of endogenous variables ζ_1 and ζ_2 were simulated to follow normal distributions with a mean of 0 and variances equal to $\psi_{11} = 1 - \gamma_{11}^2$ and $\psi_{22} = 1 - (\gamma_{21}^2 + 2\gamma_{11}\beta_{21}\gamma_{21} + \beta_{21}^2)$ so all latent variables have a variance equal to 1.

There were at least two ways to conceptualize reliability for this model. One way of considering (and calculating) reliability is on the level of the manifest variable, i.e., the reliability of the indicator *j* for latent variable *i* is equal to its common variance divided by its observed total variance which is equal to the sum of the common variance and

unique variance of the indicator $\rho_{XX} = \frac{\lambda_{Xji}^2 \Phi_{11}}{\lambda_{Xji}^2 \Phi_{11} + \theta_{\delta j}}$ where λ_{Xji} is the factor loading (the subscript *i* refers to latent variables, and the subscript *j* refers to manifest indicators), Φ_{11} is the latent factor variance, and $\theta_{\delta j}$ is the unique variance of the indicator *j*. For simulating data, reliability was manipulated at the level of indicators. All of the loadings were simulated to equal 1, and reliability of 0.7 for each manifest indicator was obtained by setting the measurement error variances to 0.4286, since $\frac{1}{(1+0.4286)} = 0.7$. Appendix B contains a more detailed description of how indicator reliability was manipulated. This choice made the assignment of accurate prior distributions for loadings more straightforward because the value of the accurate mean hyperparameter in the priors for freely estimated loadings and the value at which the loading of the first manifest variable was fixed were both 1. Had the unstandardized loadings been simulated to be equal to any value other than 1, the accurate priors for those loadings would have to be centered around the value to which the first loadings is being fixed (here, 1) and not the true value of that loading in the population. Thus, to avoid the confusion of specifying accurate priors without using the population parameter as the mean hyperparameter of the normal prior, the loadings were simulated to equal the value that the first loading for each latent variable was fixed at during estimation, i.e., 1.

Another way of considering reliability is at the level of the composite of indicators for a given latent variable. Composite reliability can be calculated using Raykov's (1997) formula for composite reliability of congeneric measures,

$\rho_{XX} = \frac{(\Sigma\lambda_i)^2 Var(\eta)}{(\Sigma\lambda_i)^2 Var(\eta) + \Sigma\theta_\varepsilon}$, where ρ_{XX} is the composite reliability, λ_i are the loadings of the indicators for a given latent variable (i = 1-3 in this model), $Var(\eta)$ is the variance of the latent variable, and $\Sigma\theta_\varepsilon$ is the sum of the measurement error variances for the three indicators of a given latent variable. The accuracy of the priors for the measurement model was manipulated at the level of composite reliability, and is described in more detail in the next section.

Independent Variables in the Simulation study

The factors manipulated in the simulation study were parameter values for structural paths, method for computing point and interval estimates/summaries of the mediated effect, and different prior specifications for Bayesian methods. The details related to each factor are presented next.

The parameter values for structural parameters in combinations 1-3 were modeled after parameter values in the paper by Finch, West, and MacKinnon (1997). The fourth parameter combination for this project was a modification of Combination 2 that was created in order to evaluate the Type I error rates of methods under examination. There are four different parameter combinations in this study, all with indicator reliability equal to 0.7, and true latent variable variances and true loadings equal to 1. The combinations were: 1) The structural parameters were γ_{11} = 0.60, β_{21} = 0.20, and γ_{21} = 0.12, which makes the direct (γ_{21} =.12) and indirect effect ($\gamma_{11}\beta_{21}$ =.12) equal; 2) The structural parameters were γ_{11} = 0.30, β_{21} = 0.40, and γ_{21} = 0.12, which makes the direct (γ_{21} =.12) and indirect effect ($\gamma_{11}\beta_{21}$ =.12) equal; 3) The structural parameters were γ_{11} = 0.30, β_{21} = 0.40, and γ_{21} = 0.36, which makes the direct (γ_{21} =.36) effect three times larger than the

indirect effect ($\gamma_{11}\beta_{21}$ =.12); and 4) The structural parameters were γ_{11} = 0, β_{21} = 0.40, and γ_{21} = 0.12, which makes the indirect effect ($\gamma_{11}\beta_{21}$) zero.

The Monte Carlo study featured both frequentist and Bayesian methods. In the frequentist analyses ML was used for point estimation, and the distribution of the product was used for interval estimation. In the Bayesian analyses the mean and median of the posterior for the mediated effect were used as point summaries, and HPD intervals were used as interval summaries. Because the distribution of the product of two random variables, such as γ_{11} and β_{21}, is not always symmetric (Craig, 1936), HPD intervals were deemed a better choice than equal-tail credibility intervals as they have the property that no value outside of the interval has higher probability than values inside of the HPD interval (Gelman, Carlin, Stern, & Rubin, 2004), which makes them better at accommodating asymmetry of the mediated effect.

Table 1

Prior specifications for Bayesian methods in the Monte Carlo Study

Informativeness		Accuracy			
		inaccurate loadings, inaccurate structural	accurate loadings, inaccurate structural	inaccurate loadings, accurate structural	accurate loadings, accurate structural
		Combination 1 (true values: $\gamma_{11} = 0.60$, $\beta_{21} = 0.20$, $\gamma_{21} = 0.12$)			
	N_{prior} = 100	$\lambda_i \sim N(.90, 84.91)$ $\gamma_{11} \sim N(.54, 80.21)$ $\beta_{21} \sim N(.13, 45.27)$ $\gamma_{21} \sim N(.05, 45.72)$	$\lambda_i \sim N(1, 84.91)$ $\gamma_{11} \sim N(.54, 80.21)$ $\beta_{21} \sim N(.13, 45.27)$ $\gamma_{21} \sim N(.05, 45.72)$	$\lambda_i \sim N(.90, 84.91)$ $\gamma_{11} \sim N(.6, 80.21)$ $\beta_{21} \sim N(.2, 45.27)$ $\gamma_{21} \sim N(.12, 45.72)$	$\lambda_i \sim N(1, 84.91)$ $\gamma_{11} \sim N(.6, 80.21)$ $\beta_{21} \sim N(.2, 45.27)$ $\gamma_{21} \sim N(.12, 45.72)$
	N_{prior} = 200	$\lambda_i \sim N(.92, 169.8)$ $\gamma_{11} \sim N(.56, 160.4)$ $\beta_{21} \sim N(.15, 90.5)$ $\gamma_{21} \sim N(.07, 91.4)$	$\lambda_i \sim N(1, 169.8)$ $\gamma_{11} \sim N(.56, 160.4)$ $\beta_{21} \sim N(.15, 90.5)$ $\gamma_{21} \sim N(.07, 91.4)$	$\lambda_i \sim N(.92, 169.8)$ $\gamma_{11} \sim N(.6, 160.4)$ $\beta_{21} \sim N(.2, 90.5)$ $\gamma_{21} \sim N(.12, 91.4)$	$\lambda_i \sim N(1, 169.8)$ $\gamma_{11} \sim N(.6, 160.4)$ $\beta_{21} \sim N(.2, 90.5)$ $\gamma_{21} \sim N(.12, 91.4)$
		Combination 2 (true values $\gamma_{11} = 0.30$, $\beta_{21} = 0.40$, $\gamma_{21} = 0.12$)			
	N_{prior} = 100	$\lambda_i \sim N(.90, 83.58)$ $\gamma_{11} \sim N(.24, 77.31)$ $\beta_{21} \sim N(.34, 72.70)$ $\gamma_{21} \sim N(.06, 79.77)$	$\lambda_i \sim N(1, 83.58)$ $\gamma_{11} \sim N(.24, 77.31)$ $\beta_{21} \sim N(.34, 72.70)$ $\gamma_{21} \sim N(.06, 79.77)$	$\lambda_i \sim N(.90, 83.58)$ $\gamma_{11} \sim N(.3, 77.31)$ $\beta_{21} \sim N(.4, 72.70)$ $\gamma_{21} \sim N(.12, 79.77)$	$\lambda_i \sim N(1, 83.58)$ $\gamma_{11} \sim N(.3, 77.31)$ $\beta_{21} \sim N(.4, 72.70)$ $\gamma_{21} \sim N(.12, 79.77)$
	N_{prior} = 200	$\lambda_i \sim N(.92, 167.2)$ $\gamma_{11} \sim N(.26, 154.6)$ $\beta_{21} \sim N(.36, 145.4)$ $\gamma_{21} \sim N(.08, 159.4)$	$\lambda_i \sim N(1, 167.2)$ $\gamma_{11} \sim N(.26, 154.6)$ $\beta_{21} \sim N(.36, 145.4)$ $\gamma_{21} \sim N(.08, 159.4)$	$\lambda_i \sim N(.92, 167.2)$ $\gamma_{11} \sim N(.3, 154.6)$ $\beta_{21} \sim N(.4, 145.4)$ $\gamma_{21} \sim N(.12, 159.4)$	$\lambda_i \sim N(1, 167.2)$ $\gamma_{11} \sim N(.3, 154.6)$ $\beta_{21} \sim N(.4, 145.4)$ $\gamma_{21} \sim N(.12, 159.4)$
		Combination 3 (true values: $\gamma_{11} = 0.30$, $\beta_{21} = 0.40$, $\gamma_{21} = 0.36$)			
	N_{prior} = 100	$\lambda_i \sim N(.90, 85.05)$ $\gamma_{11} \sim N(.24, 77.36)$ $\beta_{21} \sim N(.35, 85.37)$ $\gamma_{21} \sim N(.31, 87.22)$	$\lambda_i \sim N(1, 85.05)$ $\gamma_{11} \sim N(.24, 77.36)$ $\beta_{21} \sim N(.35, 85.37)$ $\gamma_{21} \sim N(.31, 87.22)$	$\lambda_i \sim N(.90, 85.05)$ $\gamma_{11} \sim N(.3, 77.36)$ $\beta_{21} \sim N(.4, 85.37)$ $\gamma_{21} \sim N(.36, 87.22)$	$\lambda_i \sim N(1, 85.05)$ $\gamma_{11} \sim N(.3, 77.36)$ $\beta_{21} \sim N(.4, 85.37)$ $\gamma_{21} \sim N(.36, 87.22)$
	N_{prior} = 200	$\lambda_i \sim N(.92, 170.1)$ $\gamma_{11} \sim N(.26, 154.7)$ $\beta_{21} \sim N(.36, 170.7)$ $\gamma_{21} \sim N(.32, 174.4)$	$\lambda_i \sim N(1, 170.1)$ $\gamma_{11} \sim N(.26, 154.7)$ $\beta_{21} \sim N(.26, 154.7)$ $\gamma_{21} \sim N(.32, 174.4)$	$\lambda_i \sim N(.92, 170.1)$ $\gamma_{11} \sim N(.3, 154.7)$ $\beta_{21} \sim N(.4, 154.7)$ $\gamma_{21} \sim N(.36, 174.4)$	$\lambda_i \sim N(1, 170.1)$ $\gamma_{11} \sim N(.3, 154.7)$ $\beta_{21} \sim N(.4, 154.7)$ $\gamma_{21} \sim N(.36, 174.4)$
		Combination 4	(true values: $\gamma_{11} = 0$, $\beta_{21} = 0.40$, $\gamma_{21} = 0.12$)		
	N_{prior} = 100	$\lambda_i \sim N(1.08, 82.85)$ $\gamma_{11} \sim N(.06, 76.56)$ $\beta_{21} \sim N(.46, 78.75)$ $\gamma_{21} \sim N(.17, 87.15)$	$\lambda_i \sim N(1, 82.85)$ $\gamma_{11} \sim N(.06, 76.56)$ $\beta_{21} \sim N(.46, 78.75)$ $\gamma_{21} \sim N(.17, 87.15)$	$\lambda_i \sim N(1.08, 82.85)$ $\gamma_{11} \sim N(0, 76.56)$ $\beta_{21} \sim N(.4, 78.75)$ $\gamma_{21} \sim N(.12, 87.15)$	$\lambda_i \sim N(1, 82.85)$ $\gamma_{11} \sim N(0, 76.56)$ $\beta_{21} \sim N(.4, 78.75)$ $\gamma_{21} \sim N(.12, 87.15)$
	N_{prior} = 200	$\lambda_i \sim N(1.04, 165.7)$ $\gamma_{11} \sim N(.04, 153.1)$ $\beta_{21} \sim N(.44, 157.5)$ $\gamma_{21} \sim N(.16, 174.3)$	$\lambda_i \sim N(1, 165.7)$ $\gamma_{11} \sim N(.04, 153.1)$ $\beta_{21} \sim N(.44, 157.5)$ $\gamma_{21} \sim N(.16, 174.3)$	$\lambda_i \sim N(1.04, 165.7)$ $\gamma_{11} \sim N(0, 153.1)$ $\beta_{21} \sim N(.40, 157.5)$ $\gamma_{21} \sim N(.12, 174.3)$	$\lambda_i \sim N(1, 165.7)$ $\gamma_{11} \sim N(0, 153.1)$ $\beta_{21} \sim N(.40, 157.5)$ $\gamma_{21} \sim N(.12, 174.3)$

Note. This table contains eight specifications of informative prior distributions for loadings and structural paths in the Monte Carlo study. The first hyperparameter in the normal priors is the mean, and the second hyperparameter is the precision. True values of the structural paths for each combination were included for comparison. The true value of all loadings is 1.

There were 10 different prior distributions examined in the Monte Carlo study: 2 diffuse prior specification, and 8 informative prior specifications, as shown in Table 1. The first type of prior specification in the Monte Carlo study was the briefly termed the "diffuse conjugate prior". Conjugate prior distributions lead to posterior distributions of the same parametric form (Gelman, Carlin, Stern, & Rubin, 2004). The normal distribution is a conditionally conjugate prior for loadings and structural paths in SEM, given the relevant variance terms (Kaplan & Depaoli, 2012). A fully conjugate prior density for the mean and variance parameters of a normal distribution has the product form $p(\sigma^2)p(\mu|\sigma^2)$ for which the marginal distribution of σ^2 is a scaled inverse-χ^2 (which is a special case of the inverse-gamma with $\alpha = \nu/2$ and $\beta = 1/2$, where ν denotes the degrees of freedom parameter of the inverse-χ^2) and the conditional distribution of μ given σ^2 is normal (Gelman, Carlin, Stern, & Rubin, 2004). Taking this in account and switching to the precision parametrization, the fully conjugate priors specified in this project were normal priors for loadings and structural paths centered at 0 with a precision equal to the corresponding residual precision, and gamma (G) priors with hyperparameters $\alpha = \beta = 0.5$ for measurement error precisions of manifest indicators, residual precisions of endogenous latent variables, and the precision of the exogenous latent variable. Recall that the precision is the inverse of the variance, and that the conjugate prior for the variance is an inverse-gamma, whereas the conjugate prior for the precision is a gamma distribution (Gelman, Carlin, Stern, & Rubin, 2004). Thus, when conditioning on the precision to obtain full conjugacy, the loading and structural paths are assigned normal priors conditional on the corresponding residual precisions, which are assigned gamma priors.

The second prior specification is the "generic diffuse prior" and consists of a diffuse normal prior centered at 0 with a precision hyperparameter of 0.001 for all loadings and structural paths, and a G(0.001, 0.001) as a diffuse prior for all precision parameters. Note that this is a conditionally conjugate prior where the loadings and structural paths are still assigned normal prior distributions, however, the spread of these priors is not dependent on the corresponding residual precision parameters. The mean hyperparameter of 0 for normal priors and the $\alpha = \beta = 0.001$ for the hyperparameters of the gamma prior were chosen because of the almost ubiquitous belief that such priors contribute no information to the analysis. These values were originally recommended in first programs for Bayesian analysis (Lunn, Thomas, Best, & Spiegelhalter, 2000) and are currently defaults in some software packages (Muthén, & Muthén, 1998-2015), however, there is evidence showing that gamma priors with these values of hyperparameters are not always uninformative (van Erp, Mulder, & Oberski, 2016).

The remaining eight prior distributions in the Monte Carlo study were all informative with varying degrees of informativeness and accuracy (Figure 3). In this study, informativeness was defined in terms of N_{prior}, the prior sample size used to produce the spread hyperparameter of the normal prior distributions.
Accuracy for normal priors was defined in terms of the departure of the mean hyperparameter from the simulated value for the given loading or structural path; the more different the hyperparameter from the true value, the less accurate the prior. Priors for precisions were not made inaccurate in any of the conditions.

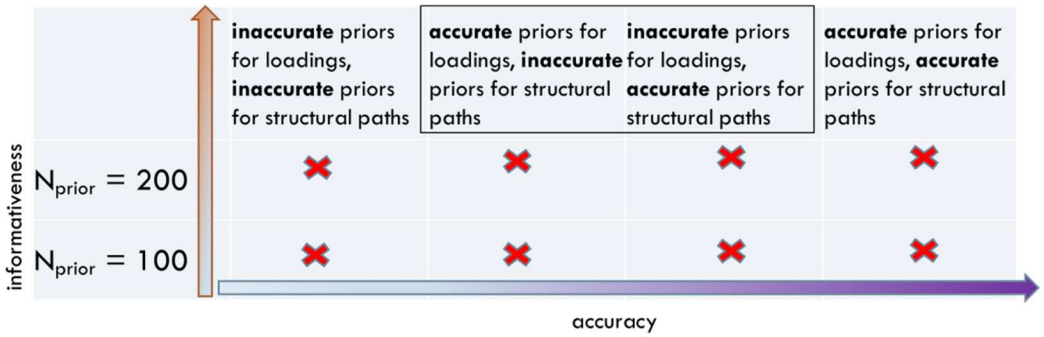

Figure 3. Eight informative prior distributions that vary levels of accuracy and informativeness. The rectangle around the conditions with partially accurate prior information, i.e., either measurement or structural parameters have accurate priors, indicates that there is currently no proof that either of the two conditions can be considered more accurate.

According to Boomsma (1982), ML estimation is robust with N = 100. Taking this information into account, as well as operating under the assumption that substantive researchers would have a hard time finding published SEM studies with N <100, the two sizes of the prior sample examined in this simulation study were N_{prior} = 100 and 200, and the size of the sample yielding the likelihood was 400.

This means that accurate and inaccurate priors were designed to carry 25% and 50% of the information carried by the likelihood. For gamma priors, the hyperparameters of accurate priors were computed analytically by plugging in the true value of the variance term and the size of the prior sample into the following formulas for the α and β hyperparameters of the gamma priors: $\theta \sim G(\nu_0 / 2, \nu_0 \sigma_0^2 / 2)$ where ν_0 is the sample size of the prior sample (N_{prior}), and σ_0^2 is the variance of the prior sample (since all priors for precisions were accurate, σ_0^2 was the true value of the corresponding variance term). Unlike the gamma distribution, the normal distribution does not have a spread parameter with a direct interpretation in terms of sample size. Normal priors that carry N_{prior} worth

of information were obtained using a series of steps. First, one sample of size N_{prior} was drawn from the population, and the Bollen-Stine (1992) transformation was used to transform the sample covariance matrix into the population (i.e., simulated) covariance matrix. The true model was then estimated using ML on the population covariance matrix. The standard errors of the structural parameters and loadings from the ML results were then used to compute the spread hyperparameters of the priors for the corresponding parameters, i.e. the value of the standard error was used as the standard deviation for the corresponding normal prior, so the precision hyperparameter for the prior being constructed was computed as the inverse of the squared standard error from the ML results. Note that the fully Bayesian version of this approach would use Bayesian methods to fit the model on the sample generating the prior, and would use standard deviations of the posteriors instead of ML standard errors to compute the spread of the priors for the main analysis. Results from ML estimation are numerically equivalent to results from a Bayesian analysis with a maximally diffuse prior, and the ML standard error is thus an approximation to the posterior standard deviation. Furthermore, point estimates and standard errors are more widely available to substantive researchers, which was another reason for choosing ML results instead of Bayesian results to create priors for the Monte Carlo study.

One of the questions that the Monte Carlo study aims to answer is whether bias in the measurement or in the structural model parameters has more severe consequences on the statistical properties of the point and interval summaries of the mediated effect. In order to directly compare the consequences of inaccurate priors in the structural and measurement parts of the model, the bias in the two sets of model parameters would have

to be equated (Cooper & Richardson, 1986). However, equating inaccuracy in the two parts of the model is a challenge for several reasons, the most important one being the different nature of the structural and measurement model parameters. It has been noted in SEM literature that "every free parameter does not contribute equally to a model's ability to fit data" (Preacher, 2006), and even commonly used fit indices have differential sensitivity to model misspecifications for factor covariances and factor loadings (Hu & Bentler, 1998). Thus, not only is there a different number of freely estimated loadings and structural paths in this model, but even if the number was the same these parameters play different roles in the model and thus it is hard to conceptualize what "equated" bias means. Furthermore, when conducting simulation studies the findings are as relevant to substantive researchers as the factors of the simulation study are representative of real data situations. Even if inaccuracy of priors for the structural and measurement models were statistically equated, this could make priors less grounded in real-life research situations, and the findings of the simulation study using such statistically equated inaccurate priors would not yield any practically useful information about these methods. For these reasons, instead of focusing on statistically equating the inaccuracy in the structural and measurement model parameters, the inaccuracy in the two sets of model parameters was made comparable by using the same mechanism to generate it. When interpreting findings, no conclusions are made about inaccuracy in one set of model parameters being more or less detrimental to the statistical properties of the mediated effect; instead, findings are discussed in terms of the way inaccuracy was induced.

The (in)accuracy of a normal prior distribution is manipulated through the mean hyperparameter, and several potential definitions of equated inaccuracy in the

measurement and structural models were considered, e.g., making the mean hyperparameters of normal priors the same number of standard deviations away from the truth, inducing the same amount of relative bias in the mean hyperparameters, and inducing the same amount of change in the chi-square statistic due to inaccurate assumptions about parameters. The definition of comparable inaccuracy chosen for this project was a prior with the mean hyperparameter that is .5 standard deviations away from the true value (note that the value of the standard deviation differs between different sizes of N_{prior}, between the loadings and structural paths, and also between different structural paths). For structural paths the mean hyperparameters of the inaccurate priors was .5 standard deviations away from the true value of the parameter. For loadings the bias of .5 standard deviations was induced in the composite reliability of a given latent variable; that is, the priors for each of the loadings had an inaccurate mean hyperparameter that when plugged into Raykov's (1997) formula for composite reliability led to the corresponding latent variable seeming like it was measured with .5 standard deviations more or less reliability compared to its reliability in the simulated model. The computation of inaccurate mean hyperparameters for the normal priors for structural parameters consists of adding/subtracting half of a standard deviation from the accurate prior, $.5sd(\gamma_{11})$, to/from the true value of the parameter, γ_{11}, in order to obtain a positively/negatively biased mean hyperparameter, γ_{11}': $\gamma_{11}' = \gamma_{11} \pm .5sd(\gamma_{11})$.

The computation of biased mean hyperparameters for loadings is slightly more complex, as the *.5sd* bias was induced in the composite reliability, and the loadings that produce this amount of relative bias were then obtained using Raykov's (1997) formula for composite reliability. Recall that he formula for the composite reliability of

congeneric measures by Raykov (1997) states that $\rho_{XX} = \frac{(\Sigma\lambda_i)^2 Var(\eta)}{(\Sigma\lambda_i)^2 Var(\eta) + \Sigma\theta_\varepsilon}$, where ρ_{XX} is the composite reliability, λ_i are the loadings of the indicators for a given latent variable (i = 1-3 in this model), $Var(\eta)$ is the variance of the latent variable, and $\Sigma\theta_\varepsilon$ is the sum of the measurement error variances for the three indicators of a given latent variable. In order to induce bias in the prior distribution for the loading, λ_i, needs to change in a way that makes ρ_{XX} have *.5sd* bias. In the simulated population all three indicators have reliability equal to 0.7 (this was accomplished by simulating the true $Var(\eta) = 1$, and the true $\theta_\varepsilon = .4286$), and the composite reliability of each latent variable is 0.8750. Since composite reliability and its standard error are not obtained in ML estimation of the model, the distribution of the composite reliability at N_{prior} = 100 and 200 had to be approximated in order to compute the standard deviation. This was done by simulating 1000 values of the two freely estimated loadings (using the true value as the mean of the normal distribution of the loading and standard deviations of .153 for N_{prior} = 200 and .216 for N_{prior} = 100), the variance of the latent variable, and measurement error variances from their true distributions based on N_{prior} = 100 and 200, plugging the values into Raykov's formula, and obtaining 1000 values of the composite reliability for the true model. The standard deviation of this empirical distribution of composite reliability became the standard deviation for the informative priors based on N_{prior} = 100 and 200. Once the standard deviation was computed, the mean hyperparameters for inaccurate priors for loadings were created by adding/subtracting *.5sd* from the true composite reliability (0.87) and calculating the values of loadings that would produce this value of

composite reliability. The calculation of the mean hyperparameter for loadings with positive bias and informativeness based on N_{prior} = 100 is illustrated next.

At N_{prior} = 100 the true composite reliability is .87, and the standard deviation of the distribution of composite reliability is 0.0407. The positively biased central tendency for the inaccurate priors for composite reliability with .5*sd* of bias in the expectation at N_{prior} = 100 is computed as .87+.5(0.0407) = 0.89. Plugging in $\rho_{XX} = 0.89$, $Var(\eta) = 1$, and $\theta_\varepsilon = .4286$ into $\rho_{XX} = \frac{(\Sigma\lambda_i)^2 Var(\eta)}{(\Sigma\lambda_i)^2 Var(\eta) + \Sigma\theta_\varepsilon}$ yields the following equation to be solved for λ_i:

$$0.89 = \frac{(\Sigma\lambda_i)^2 1}{(\Sigma\lambda_i)^2 1 + \Sigma.4286} = \frac{(3\lambda_i)^2}{(3\lambda_i)^2 + 1.2585} \Rightarrow (3\lambda_i)^2 = .89[(3\lambda_i)^2 + 1.2585]$$

$$\Rightarrow 9\lambda_i^2 = .89 \cdot 9\lambda_i^2 + .89 \cdot 1.2858 \Rightarrow (9 - .89 \cdot 9)\lambda_i^2 = 1.120 \Rightarrow 0.99\lambda_i^2 = 1.120$$

$$\lambda_i^2 = \frac{1.120}{0.99} = 1.08 \Rightarrow \lambda_i = 1.08.$$

The mean hyperparameter for inaccurate priors for loadings that produces composite reliability with .5*sd* bias at N_{prior} = 100 is 1.08. In the model fitting step in each replication the loading of the first indicator for each latent variable was set to equal the mean hyperparameter of the prior for freely estimated loadings. Thus, for conditions with accurate priors for loadings the first loading was fixed to 1, and with inaccurate priors the first loading was fixed to the inaccurate expectation for all loadings that stems from an inaccurate expectation of composite reliability.

In a given condition, the direction (sign) of the relative bias was the same for the measurement model and the structural model. For the conditions 1-3 where the mediated effect $\gamma_{11}\beta_{21}>0$ the bias in mean hyperparameter of the inaccurate priors was negative,

whereas in condition 4 where $\gamma_{11}\beta_{21}$=0 the bias was positive. This way the simulation study answers questions about changes in power with negatively biased inaccurate priors for $\gamma_{11}\beta_{21}$>0, and changes in Type I error rate with positively biased inaccurate priors for $\gamma_{11}\beta_{21}$=0. Each Bayesian analysis had 1500 burn-in iterations, 3 chains, and 5000 total iterations per chain.

Dependent Variables in the Simulation study

The statistical properties used to assess point estimates and summaries of the mediated effect are bias, relative bias (for $\gamma_{11}\beta_{21} \neq 0$), efficiency (conceptualized both as the standard error of the ML estimate and standard deviation of the posterior, and as the standard deviation of the point estimate/summary of $\gamma_{11}\beta_{21}$ over 500 replications), mean-squared error (MSE) computed as the sum of variance and the bias squared of an estimator, and root mean-squared error (RMSE) computed as $\sqrt{MSE}$.

The interval estimators of the mediated effect were evaluated in terms of Type I error rate, power, coverage, and interval width. Type I error rate and coverage were assessed according to Bradley's robustness criterion (1978). Type I error rates between 0.025 and 0.075 were deemed reasonably close to the nominal level of 0.05, and values above 0.075 were considered excessive. Coverage between 0.925 and 0.975 was considered close enough to the nominal level of 0.95, and coverage below 0.925 was deemed problematic. Imbalance was computed as the difference in the number of times the true value fell above the upper interval limit versus below the lower interval limit, and was a measure of whether a method systematically overestimates or underestimates the value of the mediated effect.

Steps in the Simulation study

The Monte Carlo study was carried out in R. A program was written to draw and store 500 samples from the four populations based on parameter values in Combinations 1-4. For each of the 500 samples per condition, ML point estimation using the package lavaan (Rosseel, 2012), interval computation using the package RMediation (Tofighi & MacKinnon, 2011), and Bayesian estimation using the packages R2WinBUGS (Sturtz, Ligges, & Gelman, 2005), coda (Plummer, Best, Cowles, & Vines, 2006), and WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000) were used to obtain point estimates, distribution of the product confidence limits, and the means, medians and HPD intervals from the posterior distributions.

In addition to point and interval estimates/ summaries of the mediated effect, the code also computed bias, relative bias for combinations with $\gamma_{11}\beta_{21}$>0, and binary indicators for power (1 if the lower limit of the confidence/credibility interval is greater than zero, and 0 if zero is in the confidence/credibility interval) for combinations where $\gamma_{11}\beta_{21}$>0, Type I error rate (1 if the confidence/credibility interval includes zero, and 0 if it doesn't) for combinations where $\gamma_{11}\beta_{21}$=0, coverage (1 if the confidence/credibility interval includes the true value, and 0 if it doesn't), and imbalance (there was two binary indicators, one if the true value is above the upper interval limit and one if the true value is below the interval limit, and imbalance is conceptualized as the difference in averages of these two binary indicators over 500 replications).

The standard error of the estimate of $\gamma_{11}\beta_{21}$ and the standard deviation of the posterior of $\gamma_{11}\beta_{21}$ were recorded at each iteration (as measures of efficiency), and so

were the width of the confidence/credibility intervals. The second way of measuring efficiency of point estimates/summaries was through the standard deviation of the point estimates/summaries of $\gamma_{11}\beta_{21}$ computed over the 500 replications. The MSE of an estimator was computed using the outputted outcome values as the sum of the variance and the bias squared of that estimator $MSE = \text{var} + bias^2$, and the RMSE was computed as $\sqrt{MSE}$. The steps in the simulation study are shown in a diagram in Figure 4. Appendix C contains sample code from the Monte Carlo study.

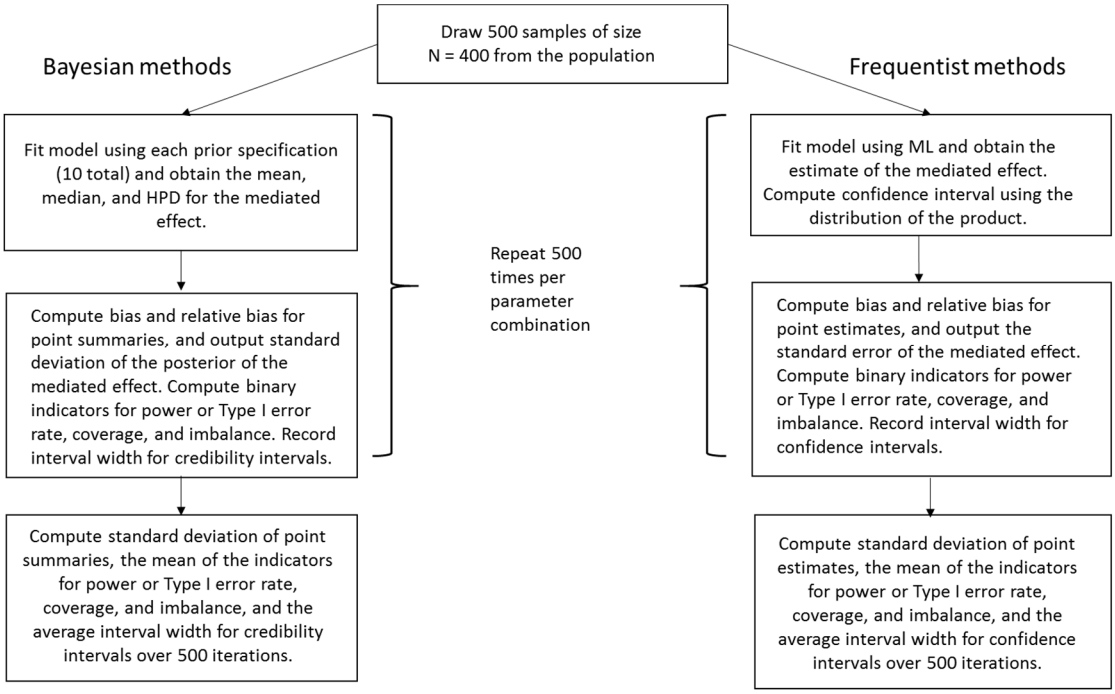

Figure 4. Diagram of steps in the Monte Carlo Study.

Results

The findings are described by hypothesis. Within each of the four sections the findings for the point estimates/ summaries are described first, followed by the results for the interval estimates/ summaries.

Bayesian methods with diffuse priors versus ML and distribution of the product

It was hypothesized that point and interval summaries obtained using Bayesian methods with diffuse priors will have comparable statistical properties to ML point estimates and interval estimates obtained using the distribution of the product. Statistical properties of point estimates/summaries are presented and described first (Table 2), followed by a display and description of statistical properties of interval estimates/summaries (Table 3).

Table 2

Statistical properties of point estimates/ summaries of the mediated effect using ML and Bayesian methods with diffuse priors

	bias	relative bias	efficiency (SE/ SD)	efficiency (SD over reps)	MSE	RMSE
ML	-.000	-.001	.046	.047	.002	.045
	.001	.008	.029	.028	.001	.032
	-.001	-.008	.027	.026	.001	.032
	.002	N/A	.023	.022	.000	.000
mean diffuse conjugate	-.000	-.001	.046	.047	.002	.045
	.001	.011	.029	.029	.001	.032
	-.001	-.006	.027	.026	.001	.032
	.002	N/A	.023	.022	.000	.000
median diffuse conjugate	-.001	-.001	-	.046	.002	.045
	-.000	-.003		.028	.001	.032
	-.002	-.017		.026	.001	.032
	.002	N/A		.021	.000	.000
mean diffuse generic	-.004	-.030	.053	.099	.010	.100
	.002	.014	.029	.029	.001	.032
	-.007	-.056	.037	.106	.011	.105
	.001	N/A	.028	.031	.001	.032
median diffuse generic	-.001	-.007	-	.047	.002	.045
	.000	.001		.029	.001	.032
	-.002	-.014		.027	.001	.032
	.002	N/A		.021	.000	.000

Note. This table contains statistical properties of point estimates of the mediated effect using ML and point summaries of the posteriors for the mediated effect using Bayesian methods with diffuse (fully) conjugate and diffuse (conditionally conjugate) "generic" priors. The four entries in each cell pertain to the 4 parameter combinations. There is no relative bias for combination 4 because the true mediated effect equals 0, and there is only one numerical value per posterior distribution for the first definition of efficiency (standard deviation of the posterior).

Findings for the point estimate of the mediated effect obtained using diffuse fully conjugate ("diffuse conjugate") and diffuse conditionally conjugate ("diffuse generic") indicate that for this set of parameter values the mean and median obtained using diffuse conjugate and the median obtained using diffuse generic priors have almost identical

bias, relative bias, efficiency, and MSE to ML estimates of the mediated effect. The mean of the posterior obtained using diffuse generic priors has higher bias and relative bias than the ML estimate and was less efficient than other point estimators/summaries in combinations 1, 3, and 4. All methods had identical RMSE values for the four combinations, except the mean of the posterior obtained using diffuse generic priors, which had a higher RMSE in combinations 1, 3, and 4 (Figure 5).

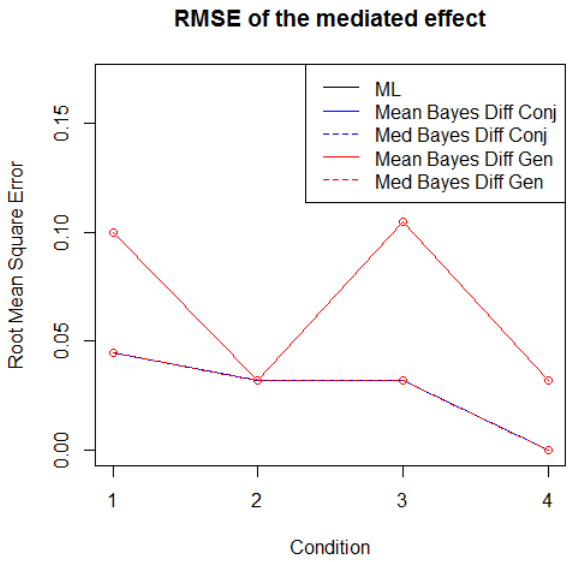

Figure 5. Plot of RMSE of the point estimates and summaries (mean and median) of the posterior for the mediated effect. The mean and median of the posterior with diffuse conjugate priors ("Diff Conj") and the median of the posterior with diffuse generic priors ("Diff Gen") have identical RMSE to ML estimates. The mean of the posterior with diffuse generic priors has the highest RMSE.

The mean is more influenced by outliers than the median, and one concern regarding this finding was that the iterations used to approximate the posterior contained some draws before the chains had converged. In order to verify if this was the case, the condition with the highest relative bias (combination 3) and diffuse generic priors was rerun with 3 chains, 15000 burn-in iterations per chain (instead of 1500), and 50000 total

iterations per chain (instead of 5000). The values of bias (-.006) and relative bias (-.053) in the posterior mean from the analyses with 15000 burn-in iterations and 50000 retained draws were almost identical between to those reported in Table 2 for those with 1500 burn-in iterations and 5000 retained draws (-.007 and -.056 for bias and relative bias, respectively). Thus the finding that the posterior mean is a more biased point summary of the mediated effect than the posterior median was not due to the mean of the posterior being computed from a mixture of draws including draws before convergence, and/or an insufficient number of MCMC total draws. A visual inspection of posterior distributions of the mediated effect obtained with diffuse generic priors revealed a positive skew, thus explaining why the posterior mean is further away from the bulk of the posterior than the posterior median. Note that the largest value of relative bias for all point summaries is still less than 2%, and that the mean of the posterior obtained using diffuse generic priors did not have excessive relative bias, i.e., relative bias above 10% as defined by Kaplan (1988), however the mean of the posterior using diffuse generic priors was a worse choice relative to the median of the same posterior, both point summaries of posteriors using diffuse conjugate priors, and the ML estimate. Thus, if deciding between ML and Bayesian methods with diffuse priors for computing point summaries for the mediated effect, it is possible to attain the same statistical properties that ML has at N = 400 by using Bayesian methods with diffuse priors, however, the choice of diffuse prior and point summary makes a small difference.

Table 3

Properties of interval estimates/ summaries of the mediated effect using the distribution of the product and Bayesian methods with diffuse priors

	power	Type I error rate	Coverage	Interval width	Imbalance
distribution of the product	.774	N/A	.948	.184	.008
	1	N/A	.960	.115	.008
	1	N/A	.966	.110	.022
	N/A	.030	.962	.092	-.022
diffuse conjugate	.770	N/A	.948	.181	.012
	1	N/A	.952	.112	.014
	1	N/A	.954	.106	.036
	N/A	.018	.974	.092	-.010
diffuse generic	.760	N/A	.952	.201	.008
	1	N/A	.956	.113	.016
	.996	N/A	.956	.132	.032
	N/A	.022	.972	.112	-.016

Note. This table contains the statistical properties of interval estimates of the mediated effect using the distribution of the product and interval summaries of the posteriors for the mediated effect using Bayesian methods with diffuse (fully) conjugate and diffuse (conditionally conjugate) "generic" priors. The four entries in each cell pertain to the 4 parameter combinations. In combination 4 the true mediated effect equals 0, so it is the only combination with values for Type I error rate instead of power.

Results for interval properties show that Bayesian HPD intervals obtained using diffuse conjugate and diffuse generic priors have comparable power and coverage values to confidence intervals obtained using the distribution of the product. Note that the findings for power come from combination 1 where the distribution of the product had power of .774, and in the remaining two combinations where the mediated effect was nonzero power for distribution of the product confidence limits was equal to 1 so there was no room for Bayesian HPD intervals to have higher power than the distribution of the product. Findings from combination 4 where the true mediated effect was zero show that Type I error rates for the distribution of the product confidence intervals were

already below 0.05, however, they were even lower for Bayesian HPD intervals with diffuse conjugate and diffuse generic priors. Interval width values are identical for distribution of the product confidence intervals and Bayesian HPD intervals with diffuse conjugate priors, but higher for HPD intervals with diffuse generic priors. Imbalance was the only outcome for which distribution of the product confidence limits had the best performance, however, this was only the case when the true mediated effect was positive. When the true mediated effect was positive, all methods tended to produce intervals for which the true value of the mediated effect was above the upper limit more often than it was below the lower limit, but confidence intervals using the distribution of the product had fewer instances of this imbalance than Bayesian HPD intervals. When the true mediated effect was zero, Bayesian HPD intervals with diffuse conjugate and diffuse generic priors had lower absolute values of imbalance than distribution of the product confidence limits.

Overall, the findings suggest that the distribution of the product confidence intervals and Bayesian HPD intervals for the mediated effect have comparable statistical properties. However, the mean of the posterior obtained using "diffuse generic" priors as a point summary of the mediated effect leads to more bias and less efficiency than ML point estimation. Thus, when using diffuse priors, it is recommended to use the median as a point summary, and to opt for "diffuse conjugate" priors when possible. These choices yield comparable statistical properties to estimates using ML and the distribution of the product, while offering the benefit of a probabilistic interpretation.

Bayesian methods with accurate informative priors versus ML and distribution of the product

It is clear without a simulation study that Bayesian methods with accurate informative priors for the structural and measurement parameters with $N_{prior} = 200$ worth of information should have the best statistical properties out of all methods in the study. However, it is not known by *how much* prior information improves bias and efficiency for point summaries, and power, Type I error rate, coverage, imbalance, and interval width for interval summaries. This study explored the amount of improvement in statistical properties of point and interval summaries of the mediated effect with the addition of accurate prior information about both measurement and structural parameters.

A comparison of statistical properties of point estimates using ML and the mean and median of posteriors obtained using accurate priors showed that Bayesian point summaries were more efficient, but ML estimates had lower bias and relative bias in the majority of cases, although not always (Table 4). This led to ML and Bayesian methods having comparable levels of MSE = 0 for the four combinations. RMSE values of Bayesian methods were slightly lower than RMSE values of ML estimates. The relative bias for all methods never exceeded 3%, but in the majority of combinations was slightly higher for Bayesian methods relative to ML. Furthermore, there was no consistent pattern of differences between the posterior mean and the median for a given combination at fixed N_{prior}; the posterior mean was less biased than the posterior median in combination 1, the median was less biased than the mean in combinations 2 and 3, and the mean and median point summaries had comparable bias in combination 4. From these results it would appear that on average, the median is less biased than the mean, however, in the

absence of a clear pattern it is difficult to make recommendations about which point summary to use to avoid bias in the mediated effect. Findings for these parameter values suggest that ML has lower bias than Bayesian methods with accurate informative priors, Bayesian methods with accurate informative priors are more efficient than ML, and if the MSE and RMSE are the only criteria then Bayesian methods with accurate informative priors are only slightly better than ML estimation.

Table 4

Properties of point estimates/ summaries of the mediated effect using ML and Bayesian methods with accurate priors

	bias	relative bias	efficiency (SE/ SD)	efficiency (SD over reps)	MSE	RMSE
ML	-.000	-.001	.046	.047	.002	.045
	.001	.008	.029	.028	.001	.032
	-.001	-.008	.027	.026	.001	.032
	.002	N/A	.023	.022	.000	.000
mean accurate N_{prior} = 100	.000	.000	.040	.035	.001	.032
	.003	.022	.025	.023	.001	.032
	.002	.014	.024	.021	.000	.000
	.002	N/A	.020	.018	.000	.000
median accurate N_{prior} = 100	-.001	-.005	-	.035	.001	.032
	.001	.012		.023	.001	.032
	.001	.004		.021	.000	.000
	.002	N/A		.018	.000	.000
mean accurate N_{prior} = 200	-.000	-.002	.036	.028	.001	.032
	.003	.023	.023	.019	.000	.000
	.002	.016	.022	.018	.000	.000
	.001	N/A	.018	.016	.000	.000
median accurate N_{prior} = 200	-.001	-.008	-	.028	.001	.032
	.002	.014		.019	.000	.000
	.001	.007		.018	.000	.000
	.001	N/A		.015	.000	.000

Note. This table contains statistical properties of point estimates of the mediated effect using ML and point summaries of the posteriors for the mediated effect using Bayesian methods with accurate priors based on N_{prior} = 100 and 200. The four entries in each cell pertain to the 4 parameter combinations. There is no relative bias for combination 4 because the true mediated effect equals 0, and there is only one numerical value per posterior distribution for efficiency defined as the average standard error of the mediated effect or average standard deviation of the posterior for the mediated effect.

As expected, Bayesian HPD intervals obtained using accurate priors had more power, lower Type I error rates, higher coverage, lower interval width, and less imbalance than distribution of the product confidence intervals (Table 5). As in the diffuse case, there was no way of evaluating whether Bayesian methods with accurate

priors have higher power than ML in combinations 2 and 3 because power using ML was equal to 1. Thus the finding that Bayesian methods can increase power by .132 (17%) when the accurate prior has the weight of 25% of the likelihood, and by .192 (25%) when the accurate prior has the weight of 50% of the likelihood is based on only one parameter combination. Also, it is worth noting that whereas the Type I error rates and coverage of the distribution of the product confidence limits were within Bradley's (1978) robustness criterion (i.e., between 0.025 and 0.075 for Type I error rate, and between 0.925 and 0.975 for coverage), the Type I error rates of HPDs with accurate priors was lower than the robustness criterion (i.e., below 0.025) and coverage was above the robustness criterion (i.e., above 0.975) in three out of the four parameter combinations. The decreases in interval width for HPDs relative to distribution of the confidence limits were between 13-15% when the accurate prior had 25% of the weight of the likelihood, and between 20-24% when the accurate prior had 50% of the weight of the likelihood.

Table 5

Properties of interval estimates/ summaries of the mediated effect using the distribution of the product and Bayesian methods with accurate priors

	power	Type I error rate	Coverage	Interval width	Imbalance
distribution of the product	.774	N/A	.948	.184	.008
	1	N/A	.960	.115	.008
	1	N/A	.966	.110	.022
	N/A	.03	.962	.092	-.022
accurate $N_{prior} = 100$	.906	N/A	.972	.156	.004
	1	N/A	.968	.099	.000
	1	N/A	.980	.095	.012
	N/ A	.008	.990	.080	-.006
accurate $N_{prior} = 200$	.966	N/A	.972	.140	.004
	1	N/A	.982	.090	-.002
	1	N/A	.984	.087	.012
	N/ A	.008	.990	.072	-.006

Note. This table contains the statistical properties of interval estimates of the mediated effect using the distribution of the product and interval summaries of the posteriors for the mediated effect using Bayesian methods with accurate priors based on $N_{prior} = 100$ and 200. The four entries in each cell pertain to the 4 combinations. Note that only combination 4 has a mediated effect equal to 0, thus it is the only combination with values for Type I error rate instead of power.

Findings suggest that using Bayesian methods with accurate informative priors led to slightly more bias and higher efficiency of point summaries of the posterior for the mediated effect relative to ML estimates of the mediated effect. Bayesian HPD intervals with accurate priors had Type I error rates closer to 0, power and coverage closer to 1, interval width that was at least 13% lower, and less imbalance relative to distribution of the product confidence limits. An examination of bias in the posterior mean and median for a given parameter combination showed that the mean was less biased then the median in combination 1, in combinations 2 and 3 the median was less biased than the mean, and in comb 4 they were equivalent. From these results it would appear that the median is might be less biased than the mean, however, more parameter combinations would need

to be studied in order to make any recommendations about which point summary is less biased with accurate informative priors.

Bayesian methods with inaccurate informative priors versus ML and distribution of the product

Conversely, it appears intuitive that Bayesian methods with inaccurate informative priors for the structural and measurement parameters with N_{prior} = 200 worth of information would have the worst statistical properties out of all the methods in the study. It is also unclear how bad the statistical properties of the mediated effect are with N_{prior} = 100 and 200 with inaccurate priors for the measurement and/or structural parameters. This study quantified the risks of using inaccurate informative priors that have the weight of a quarter and a half of the likelihood.

Similar to the accurate prior case, the comparison of ML to Bayesian methods with inaccurate priors suggests that ML point estimates had less bias and relative bias, but the Bayesian methods were more efficient (Table 6). However, even with inaccurate priors that carried half of the weight of the likelihood, the relative bias for Bayesian point summaries did not exceed 5% for any of the parameter combinations. If MSE and RMSE are the only criteria for selecting between the two methods, then Bayesian methods with inaccurate priors would be a better choice than ML in most parameter combinations. However, the range of differences between RMSE of ML and Bayesian methods in Table 6 is between 0 and .032, and in the absence of guidelines about differences in RMSE that make on method superior to another, such differences in RMSE could be considered negligible, thus implying that the two methods have comparable bias and efficiency.

Table 6

Statistical properties of point estimates/ summaries of the mediated effect using ML and Bayesian methods with inaccurate priors for structural and measurement parameters

	bias	relative bias	efficiency (SE/ SD)	efficiency (SD over reps)	MSE	RMSE
ML	-.000	-.001	.046	.047	.002	.045
	.001	.008	.029	.028	.001	.032
	-.001	-.008	.027	.026	.001	.032
	.002	N/A	.023	.022	.000	.000
mean inaccurate N_{prior} = 100	-.003	-.023	.040	.035	.001	.032
	-.001	-.012	.025	.023	.001	.032
	-.003	-.021	.024	.022	.000	.000
	.006	N/A	.021	.018	.000	.000
median inaccurate N_{prior} = 100	-.003	-.028	-	.035	.001	.032
	-.003	-.022		.023	.001	.032
	-.004	-.030		.022	.000	.000
	.006	N/A		.018	.000	.000
mean inaccurate N_{prior} = 200	-.004	-.036	.036	.029	.001	.032
	-.002	-.020	.023	.019	.000	.000
	-.003	-.027	.022	.018	.000	.000
	.006	N/A	.019	.016	.000	.000
median inaccurate N_{prior} = 200	-.005	-.042	-	.029	.001	.032
	-.003	-.029		.019	.000	.000
	-.004	-.035		.018	.000	.000
	.006	N/A		.016	.000	.000

Note. This table contains statistical properties of point estimates of the mediated effect using ML and point summaries of the posteriors for the mediated effect using Bayesian methods with inaccurate priors for both measurement and structural parameters based on N_{prior} = 100 and 200. The four entries in each cell pertain to the parameter 4 combinations. There is no relative bias for combination 4 because the true mediated effect equals 0, and there is only one numerical value per posterior distribution for the first definition of efficiency.

Even though the inaccurate priors in combinations 1-3 that had a positive true mediated effect had negatively biased mean hyperparameters, Bayesian HPDs still had higher power than the distribution of the product confidence limits (Table 7). Also, in combination 4 where the true mediated effect was zero, the Type I error rates became

lower with more informativeness, despite the fact that the inaccurate priors in this combination were positively biased. Thus, it appears that .5*sd* of negative bias did not produce lower power, nor did .5*sd* of positive bias produce excessive Type I error rates.

Table 7

Properties of interval estimates/ summaries of the mediated effect using the distribution of the product and Bayesian methods with inaccurate priors for structural and measurement parameters

	power	Type I error rate	Coverage	Interval width	Imbalance
distribution of the product	.774	N/A	.948	.184	.008
	1	N/A	.960	.115	.008
	1	N/A	.966	.110	.022
	N/A	.03	.962	.092	-.022
inaccurate both $N_{prior} = 100$	.890	N/A	.960	.155	.012
	1	N/A	.954	.097	.018
	1	N/A	.968	.093	.022
	N/A	.028	.972	.082	-.028
inaccurate both $N_{prior} = 200$	.952	N/A	.976	.139	.012
	1	N/A	.968	.088	.020
	1	N/A	.978	.085	.018
	N/A	.018	.982	.074	-.018

Note. This table contains statistical properties of interval estimates of the mediated effect using the distribution of the product and interval summaries of the posteriors for the mediated effect using Bayesian methods with inaccurate priors for both measurement and structural parameters based on $N_{prior} = 100$ and 200. The four entries in each cell pertain to the 4 parameter combinations. The true mediated effect in combination 4 equals 0, thus it is the only combination with values for Type I error rate instead of power.

Coverage for Bayesian methods with inaccurate priors was within Bradley's robustness criterion (1978), i.e., between .925 and .975, which is lower than coverage for Bayesian methods with accurate priors and the same levels of informativeness. Thus, the inaccuracy in the priors seems to have reduced the tendency Bayesian methods have that with more informative accurate priors coverage goes to 1. Interval width was lower for

Bayesian methods with informative priors than for distribution of the product confidence limits, but imbalance tended to be higher in the majority of combinations.

The findings suggest that inaccuracy in the mean hyperparameters of informative priors has undesirable effects in terms of bias and relative bias for point summaries and imbalance (and potentially coverage) for interval summaries of the mediated effect, however, even with inaccurate priors that have 50% of the weight of the likelihood the relative bias remained below 5% and imbalance was below 4%. Despite the inaccuracy of the mean hyperparameters, more informativeness in the prior decreased interval width and led to Type I error rates tending toward zero, and power tending toward 1.

Information in the priors for structural paths versus loadings

It is expected that a .5*sd* inaccuracy in the expectation for structural paths will have more bearing on the statistical properties of the point and interval summaries of the mediated effect than a .5*sd* inaccuracy in the expectation for composite reliability. A comparison of statistical properties of the point summaries for the mediated effect at a given size of N_{prior} partially supported the hypothesis that inaccurate priors for structural parameter lead to worse statistical properties than inaccurate priors for measurement model parameters when the prior expectations for structural parameters and composite reliability are .5*sd* away from the simulated values (Table 8). At a fixed value of N_{prior} and for a given point summary (mean or median), the amount of bias and relative bias was larger if the inaccurate priors were assigned to structural than to measurement model parameters.

Table 8

Properties of point estimates/ summaries of the mediated effect using ML and Bayesian methods with inaccurate priors

	bias	relative bias	efficiency (SE/ SD)	efficiency (SD over reps)	MSE	RMSE
ML	-.000	-.001	.046	.047	.002	.045
	.001	.008	.029	.028	.001	.032
	-.001	-.008	.027	.026	.001	.032
	.002	N/A	.023	.022	.000	.000
$N_{prior} = 100$						
mean inaccurate measurement	.002	.021	.040	.036	.001	.032
	.005	.042	.026	.023	.001	.032
	.004	.032	.024	.022	.000	.000
	.002	N/A	.020	.018	.000	.000
median inaccurate measurement	.002	.016	-	.036	.001	.032
	.004	.032		.023	.001	.032
	.003	.023		.022	.000	.000
	.001	N/A		.017	.000	.000
mean inaccurate structural	-.005	-.044	.039	.034	.001	.032
	-.004	-.032	.025	.022	.001	.032
	-.005	-.040	.024	.021	.000	.000
	.006	N/A	.021	.019	.000	.000
median inaccurate structural	-.006	-.048	-	.034	.001	.032
	-.005	-.042		.022	.001	.032
	-.006	-.049		.021	.000	.000
	.006	N/A		.018	.000	.000
$N_{prior} = 200$						
mean inaccurate measurement	.002	.019	.036	.030	.001	.032
	.005	.045	.023	.020	.000	.000
	.004	.036	.022	.019	.000	.000
	.001	N/A	.018	.015	.000	.000
median inaccurate measurement	.002	.013	-	.029	.001	.032
	.004	.035		.020	.000	.000
	.003	.028		.019	.000	.000
	.001	N/A		.015	.000	.000
mean inaccurate structural	-.007	-.058	.035	.028	.001	.032
	-.005	-.042	.023	.019	.000	.000
	-.006	-.048	.022	.018	.000	.000
	.006	N/A	.019	.017	.000	.000
median inaccurate structural	-.008	-.063	-	.028	.001	.032
	-.006	-.051		.019	.000	.000
	-.007	-.056		.018	.000	.000
	.006	N/A		.016	.000	.000

Note. Table 8 contains statistical properties of point estimates of the mediated effect using ML and point summaries of the posteriors for the mediated effect using Bayesian methods with inaccurate priors for measurement or structural parameters based on $N_{prior} = 100$ and 200. The four entries in each cell pertain to the 4 parameter combinations. There is no relative bias for combination 4 because the true mediated effect equals 0, and there is only one numerical value per posterior for the first definition of efficiency (standard deviation of the posterior).

However, the direction (sign) of the bias was different too. Inaccurate priors in the measurement model led to positive bias for all four combinations, even in combination 4 where the bias in the priors was negative. Inaccurate priors in the structural model always led to bias in the same direction as the bias of the mean hyperparameter of the normal prior. Figure 6 shows values of bias of the mean and median of the posterior for the mediated effect as a function of the sign of the bias in the mean hyperparameter of the normal priors for structural paths and loadings. Black dots represent combinations where only measurement model parameters had inaccurate priors, and red dots represent combinations where only structural parameters had inaccurate priors.

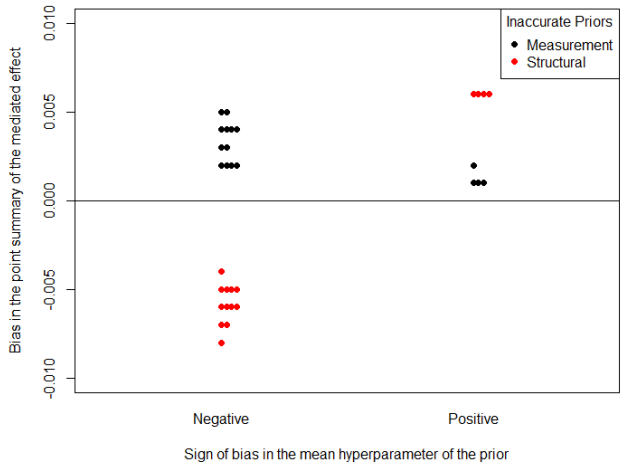

Figure 6. Bias due to inaccurate priors for structural and measurement model parameters. This plot contains values of bias in the point summaries (mean and median) of the mediated effect at N_{prior} = 100 and 200 as a function of the sign of the bias in the mean hyperparameter of the normal priors for loadings ("Measurement", black dots) and structural ("Structural", red dots) parameters.

Recall that in combinations 1-3 the bias in the normal priors was negative in order to evaluate the reduction in power with negatively biased prior expectations for structural parameters and composite reliability. In combination 4 the bias in the normal priors was

positive in order to evaluate the increase in Type I error rates with positive bias in the prior expectations about structural parameters and composite reliability. With negatively biased prior expectations for structural parameters, the bias in the point summaries for the mediated effect is also negative. Conversely, with positively biased prior expectations for structural parameters, the bias in the mediated effect is positive. However, regardless of whether the expectation for composite reliability (i.e. measurement model parameters) is positively or negatively biased, the resulting bias in point summaries for the mediated effect is always positive. For example, recall that inaccurate priors in combinations 1-3 have negatively biased mean hyperparameters that underestimate composite reliability and underestimate values of structural paths; looking at the part of table 8 pertaining to combination 1 at $N_{prior} = 200$, the bias for the posterior mean of the mediated effect when the measurement model parameters have inaccurate priors is positive (0.002), and the bias for the posterior mean of the mediated effect when the structural parameters have inaccurate priors is negative (-0.007). In other words, with negative bias in the inaccurate priors (conditions 1-3), point summaries obtained using inaccurate priors for the measurement model have positive bias (i.e. bias in the *opposite* direction than bias in the prior), and point summaries obtained using inaccurate priors for the structural model have negative bias (i.e., bias in the *same* direction as bias in the prior).

The biggest difference in absolute value of relative bias produced by inaccurate priors in the measurement versus structural parameters was 5%, and even for this parameter combination all point summaries had relative bias below 10%. Thus, .5*sd* of bias in the hyperparameters of priors for measurement and structural model parameters does produce more bias if the inaccurate priors are for structural instead of measurement

model parameters. Point summaries following estimation with inaccurate priors for the measurement versus structural parameters were equally efficient, and even though Bayesian methods had lower values of RMSE, the difference in RMSE between ML and Bayesian methods with inaccurate priors was at most .032. Bayesian point summaries with inaccurate priors were more biased and more efficient than ML estimates.

Findings for point summaries partially support the hypothesis that .5*sd* of bias has more severe consequences for prior expectations for structural parameters than prior expectations about composite reliability (i.e., measurement model parameters). That is, inaccurate priors for structural parameters led to more bias and relative bias than inaccurate priors for loadings, however the efficiency of Bayesian point summaries remained comparable regardless of which part of the model was assigned inaccurate priors. Another important finding that emerged is that inaccurate priors for loadings led to positive bias regardless of the sign of the bias in the hyperparameter of the prior, whereas the direction of the bias in the point summary following estimation with inaccurate priors for structural parameters is the same as the direction of bias in the mean hyperparameter of the inaccurate prior for the structural path.

The hypothesis that structural bias is more detrimental than measurement bias for interval summaries of the mediated effect was partially supported. Inaccurate priors for structural parameters led to lower power in combination 1 (in combinations 2 and 3 all methods had power of 1) and more imbalance than inaccurate priors for measurement parameters (Table 9). The inaccurate priors of .5*sd* did not produce any instances of power below 0.8 or of Type I error rates above 0.05; recall that the direction of the inaccuracy in the mean hyperparameters was designed to decrease power in combinations

with the true mediated effect above 0, and to increase Type I error rate in the combination with the true mediated effect equal to zero. The inaccuracy of priors for the *structural* parameters led to Type I error rates and coverage either within or closer to limits Bradley's robustness criterion. The Type I error rates with inaccurate priors for *measurement* parameters was practically zero, and the coverage was closer to 1 than when priors were inaccurate for the structural parameters. Thus, the inaccuracy in the structural model combated the tendency of Bayesian methods with accurate informative priors to have Type I error rates of 0 and coverage of 1. The inaccuracy in the measurement model did not change this tendency. Bayesian HPDs with inaccurate priors for structural parameters had lower interval width than HPDs with inaccurate priors for measurement parameters. Finally, even with inaccurate priors for either measurement or structural parameters, Bayesian HPDs had higher power and lower interval width than the distribution of the product confidence limits.

Table 9

Properties of interval estimates/ summaries of the mediated effect using the distribution of the product and Bayesian methods with inaccurate priors

	power	Type I error rate	Coverage	Interval width	Imbalance
distribution of the product	.774	N/A	.948	.184	.008
	1	N/A	.960	.115	.008
	1	N/A	.966	.110	.022
	N/A	.03	.962	.092	-.022
			$N_{prior} = 100$		
inaccurate measurement	.904	N/A	.966	.157	-.002
	1	N/A	.968	.099	.000
	1	N/A	.974	.095	.006
	N/A	.010	.988	.080	-.008
inaccurate structural	.892	N/A	.964	.153	.014
	1	N/A	.956	.096	.028
	1	N/A	.962	.092	.034
	N/A	.032	.966	.082	-.030
			$N_{prior} = 200$		
inaccurate measurement	.966	N/A	.974	.142	.002
	1	N/A	.978	.091	-.010
	1	N/A	.978	.087	.006
	N/A	.004	.994	.072	-.002
inaccurate structural	.950	N/A	.978	.138	.018
	1	N/A	.964	.087	.028
	1	N/A	.976	.084	.024
	N/A	.028	.972	.074	-.028

Note. This table contains statistical properties of interval estimates of the mediated effect using the distribution of the product and interval summaries of the posteriors for the mediated effect using Bayesian methods with inaccurate priors for measurement or structural parameters based on N_{prior} = 100 and 200. The four entries in each cell pertain to the 4 combinations. In combination 4 the true mediated effect equals 0, thus it is the only combination with values for Type I error rate instead of power.

Overall findings suggest that .5*sd* of inaccuracy is more detrimental when it occurs in the prior expectations for structural parameters than in prior expectations for composite reliability: point summaries have more bias and relative bias, power is lower, and imbalance increases. Bias in priors for structural parameters led to lower coverage

and higher Type I error rate than bias in the loadings; however, for the amount and type of bias tested in this study this decrease did not produce Type I error rate above 0.05 nor coverage below 0.95. Another important findings is that the direction of bias in the mediated effect reflects the direction of the bias in the mean hyperparameters of the *structural* parameters, and the bias in the mediated effect was positive regardless of the direction of inaccuracy in the expectation about *measurement* parameters.

CHAPTER 3

DISCUSSION

The Monte Carlo study was designed to answer four questions about the statistical properties of the mediated effect in the single mediator model with latent variables: what are the risks and benefits of switching to the Bayesian framework for computing the mediated effect, what are the benefits in the best case scenario in this study when priors are accurate and most informative, what are the risks in the worst case scenario when priors are innaccurate and most informative, and does the same amount of bias in the expectations about *structural* or *measurement* model parameters have more bearing on the statistical properties of the point and interval summaries of the mediated effect?

Bayesian methods with diffuse fully conjugate priors had statistical properties that were as good as those using ML estimation, but switching to the Bayesian framework gives the benefit of probabilistic interpretations. However, it is not possible to specify fully conjugate priors in all software packages. Findings from this study show that when using diffuse generic priors (as defined in this study), it is better to use the posterior median instead of the posterior mean as the point summary for the mediated effect.

In the best case scenario tested in this study, point summaries from Bayesian methods with accurate priors that carry 25% and 50% of the weight of the likelihood were more biased and more efficient than ML estimates, and interval summaries were at least 13% and up to 24% narrower than distribution of the product confidence limits. The increases in power with the use of accurate informative priors described in this study ranged between .132 (17%) and .192 (25%). Type I error rates tended toward zero while coverage tended toward 1 with increases in the amount of accurate prior information. The

tendency of Bayesian credibility intervals with accurate informative priors to have Type I error rates equal to zero and coverage equal to 1 also occurs in the single mediator model with manifest variables (Miočević, MacKinnon, & Levy, 2016). Even though Type I error rates below 0.025 and coverage above 0.975 are outside of what this study defines as nominal values, this is not necessarily a limitation of Bayesian methods. It is easy to understand why Type I error rates above 0.075 and coverage below 0.925 are a problem, however, it is harder to decide whether lower than nominal Type I error rates and overcoverage are undesirable. If one considers it important that the method has nominal values of criteria, then Type I error rates below nominal values and overcoverage are a problem, however, if the goal is to minimize false positives and produce accurate intervals for a parameter, then this situation is more desirable than having empirical Type I error rates and coverage equal to nominal values.

The hypothesis that Bayesian methods with the most informative inaccurate priors would have the worst statistical properties was only partly supported: bias and relative bias increased with inaccurate priors, but relative bias was never above 4.2%, and HPD intervals had higher imbalance than distribution of the product confidence limits. Somewhat surprisingly, even with inaccurate priors, Bayesian methods had higher power than ML estimation, and satisfactory Type I error rates and coverage. This finding makes sense given that even the inaccurate priors in conditions where power was tested had positive mean hyperparameters. Accurate and inaccurate priors for the mediated effect, computed as the product of draws from priors for γ_{11} and β_{21}, were visually inspected for combination 1 at $N_{prior} = 100$ and $N_{prior} = 200$ (Figure 7). The goal was to evaluate the difference these priors place on negative and zero values of $\gamma_{11}\beta_{21}$. At $N_{prior} = 100$, the

induced accurate prior for $\gamma_{11}\beta_{21}$ places 10% of its weight on the area $\gamma_{11}\beta_{21.} \leq 0$, and the inaccurate prior places 20% on the same area. At $N_{prior} = 200$ the accurate prior for $\gamma_{11}\beta_{21}$ places 2% on the area $\gamma_{11}\beta_{21.} \leq 0$, and the inaccurate prior places 9% of its weight on the same area. Even though the "inaccurate prior" is inaccurate in the sense that the central tendency is not correct, it still places most of the density above 0, so it's actually correct in saying that the mediated effect is most likely positive. Thus, one conclusion that emerged from this finding is that the accuracy of a prior distribution depends on the inferential goals of the analysis. In this case, the goals were to test the bias of the point summary of the mediated effect, and to evaluate power (i.e., test whether the 0 is in the credibility interval for the mediated effect) with inaccurate priors. The "inaccurate priors" in this study were inaccurate for the purposes of estimating the value of the mediated effect without bias, but not for the purposes of testing whether the credibility interval for the mediated effect contains 0.

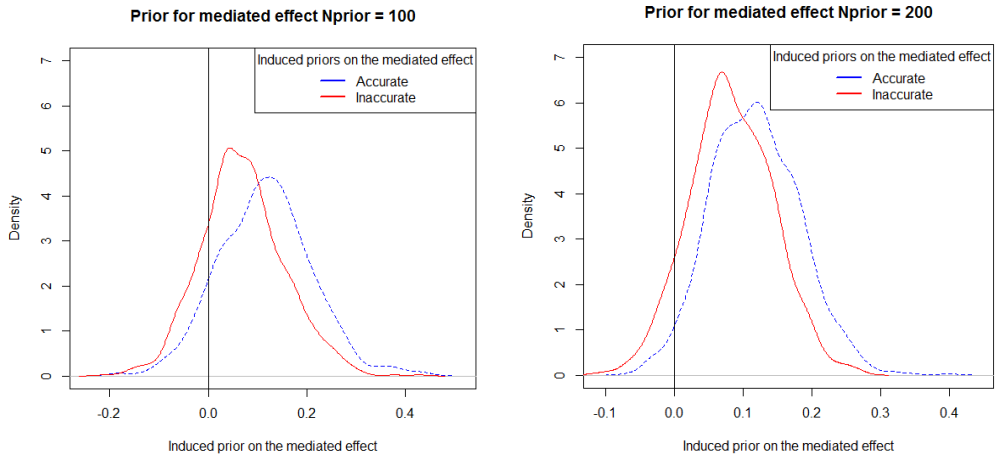

Figure 7. Induced accurate and inaccurate priors for the mediated effect. Induced priors on the mediated effect for combination 1 were computed from 1000 simulated draws from accurate and inaccurate priors for paths γ_{11} and β_{21}. Vertical lines indicate $\gamma_{11}\beta_{21} = 0$, and the difference in the area to the left of the line for the blue and the red lines indicates the difference in the weight these priors place on values of the mediated effect less than and equal to zero.

When looking at the statistical properties of the mediated effect with accurate and inaccurate priors holding condition and N_{prior} constant, it seems that certain statistical properties, such as efficiency for point estimates/summaries (Figure 9) and power (Figure 10) and interval width (Figure 12) for interval estimates/summaries, are not as influenced by the accuracy of the priors (at least with the levels of inaccuracy tested in this study) as they are influenced by their informativeness. Relative bias (Figure 8) and imbalance (Figure 13), on the other hand, appear to be more dependent on the (in)accuracy in the hyperparameters for normal priors, and Type I error rates (Figure 10, and Tables 5 and 7) and coverage (Figure 11) seem to be influenced by both accuracy and informativeness. It is important to emphasize that all conclusions from the Monte Carlo study hold only for the parameter values and prior specifications used in this study, and that the statistical properties of posterior summaries produced with a given inaccuracy in the prior are dependent on the informativeness, and the statistical properties of posterior summaries produced with a given informativeness in the prior are dependent on the accuracy. In other words, accuracy and informativeness were considered separately when constructing the priors, however, their impact on the statistical properties of the posterior summaries cannot be disentangled.

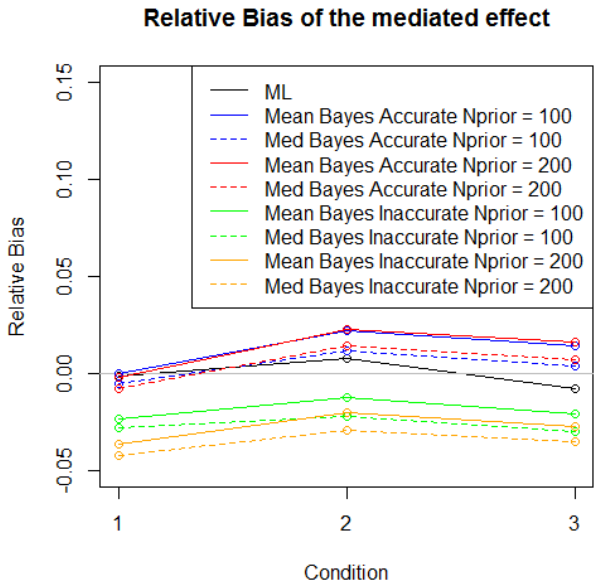

Figure 8. Relative bias of point summaries of the mediated effect with accurate and inaccurate priors. DP refers to the distribution of the product. Inaccurate priors led to point summaries with higher absolute relative bias than accurate priors.

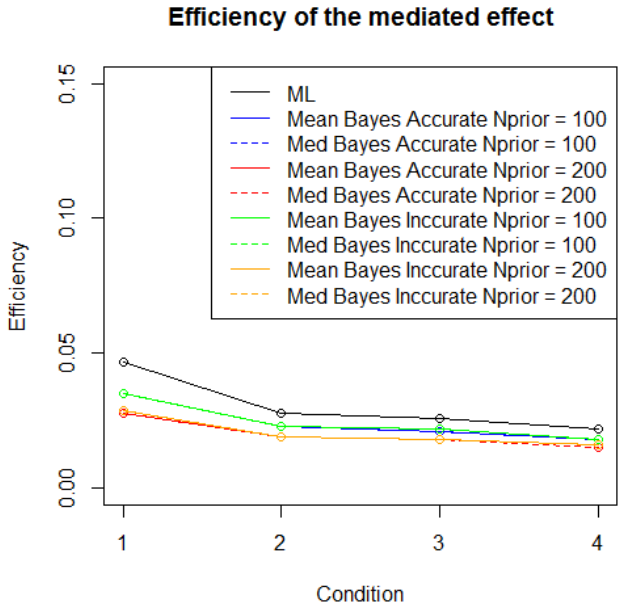

Figure 9. Efficiency of point summaries of the mediated effect with accurate and inaccurate priors. DP refers to the distribution of the product. More informativeness in the prior led to lower efficiency.

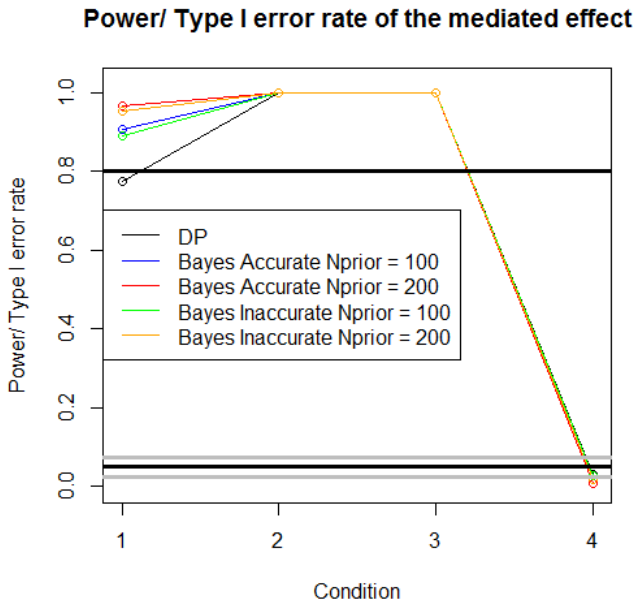

Figure 10. Power and Type I error rates of interval summaries for the mediated effect with accurate and inaccurate priors. DP refers to the distribution of the product. More informativeness in the prior led to more power, and more accuracy and informativeness in the prior led to lower Type I error rates.

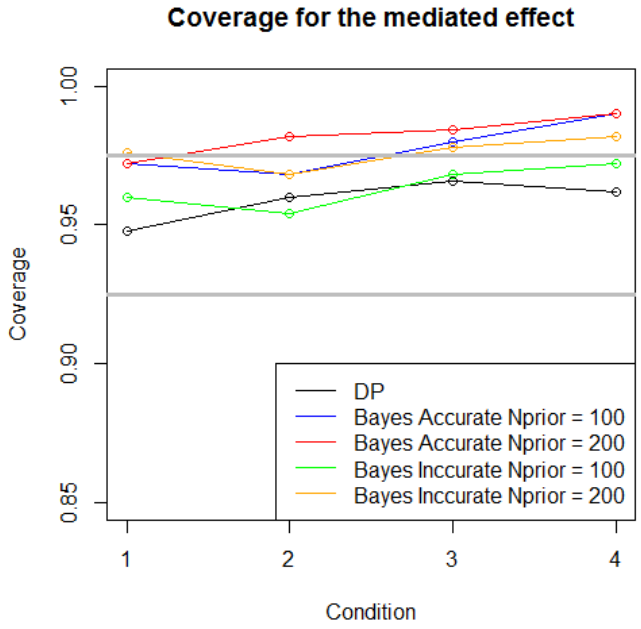

Figure 11. Coverage of interval summaries for the mediated effect with accurate and inaccurate priors. DP refers to the distribution of the product. Higher accuracy and more informativeness led to higher coverage.

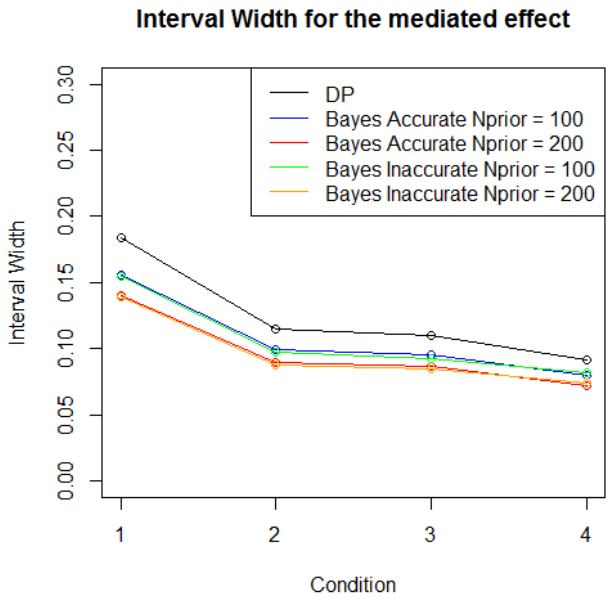

Figure 12. Interval width of interval summaries of the mediated effect with accurate and inaccurate priors. DP refers to the distribution of the product. More informativeness led to lower interval width.

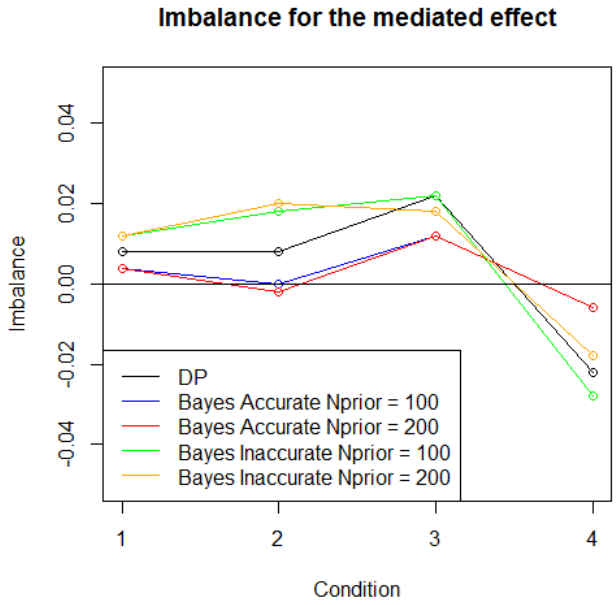

Figure 13. Imbalance of interval summaries of the mediated effect with accurate and inaccurate priors. DP refers to the distribution of the product. More accuracy led to less imbalance.

Finally, the statistical properties of the mediated effect were more affected by .5 standard deviations of inaccuracy in the expectations about structural parameters than by the same amount of inaccuracy in the expectations about composite reliability of manifest indicators. The direction of the bias in the point summaries and of power, Type I error rate, and imbalance of the mediated effect followed (i.e. was the same as) the direction of the inaccuracy in the expectations about *structural* parameters, but did not follow the direction of the inaccuracy in the expectations about *measurement* parameters. In order to examine this finding further, the bias in the posterior summaries for the loadings was also inspected, and it follows the direction of the bias in the prior for loadings. It is unclear why the point summaries of the mediated effect have positive bias when the point summaries of the loadings have both positive and negative bias due to inaccurate priors. It is possible that the inaccurate priors for loadings induce bias in the posterior covariances between latent variables, and the bias in the mediated effect is due to the fact that the mediated effect are a function of these covariances. It would be important to follow up on this unusual finding and examine the mechanism behind it and find out whether this finding is specific to parameter combinations tested in this study or if it is a more general occurrence.

One limitation of this study is that statistical properties of ML estimation were already fairly satisfactory at N = 400, thus there may not have been enough "room" for improvement to occur with Bayesian methods with accurate informative priors. Furthermore, the finding that inaccurate informative priors do not lead to the worst performance for all statistical properties of the mediated effect may not generalize to

situations with more than .5*sd* of inaccuracy in the expectations about structural paths and composite reliability.

Some future directions for this line of research are to examine the benefits and risks of using accurate and inaccurate informative prior distributions in the same model with samples smaller than 400. Furthermore, it is important for the applicability of methodological research to consider different ways bias occurs in the measurement and structural models and to describe how to design factors for simulation studies that closely match the types of inaccurate priors that might occur in practice. Finally, some ideas that emerged during the writing of this dissertation, but were not pursued immediately, are to examine statistical properties of Bayesian methods for SEM with nonnormal indicators, and to explore ways of creating prior information from mediation models with manifest variables for Bayesian analyses of mediation models with latent variables. Bayesian SEM is an active area of research, and as of now, there are still no clear guidelines for substantive researchers about how prior information ought to be used in Bayesian estimation of these models without biasing the results. This dissertation is the beginning of a line of research aiming to create such guidelines.

REFERENCES

Aitkin, M., & Aitkin, I. (2005). Bayesian inference for factor scores. *Contemporary psychometrics: A festschrift for Roderick P. McDonald*, 207-222.

Arbuckle, J. (2007). *Amos 16.0 user's guide*. Chicago, IL: SPSS.

Asparouhov, T., & Muthén, B. (2010). Bayesian analysis of latent variable models using Mplus. *Retrieved from http://www.statmodel2.com/download/BayesAdvantages18.pdf on June*, *17*, 2014.

Barnard, J., McCulloch, R. E., & Meng, X. L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica 10*, 1281-1311.

Bartlett, M. S. (1957). A Comment on D. V. Lindley's Statistical Paradox. *Biometrika*, *44*, 533-534. doi: https://doi.org/10.1093/biomet/44.3-4.533

Bayarri, M. J., & Berger, J. O. (2000a). P values for composite null models. *Journal of the American Statistical Association, 95*, 1127–1142. doi: http://dx.doi.org/10.1080/01621459.2000.10474309

Brooks, S. P. (1998). Markov chain Monte Carlo method and its application. *The Statistician, 47,* 69-100. doi: https://doi.org/10.1111/1467-9884.00117

Brooks, S. P., & Roberts, G. O. (1998). Convergence assessment techniques for Markov chain Monte Carlo. *Statistics and Computing, 8*(4)*,* 319-335. doi: https://doi.org/10.1023/a:1008820505350

Casella, G., & George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician, 46,* 167-174. doi: https://doi.org/10.2307/2685208

Celeux, G., Forbers, F., Robert, C., & Titterington, D. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, 1, 651–674. doi: https://doi.org/10.1214/06-ba122

Chen, J., Choi, J., Weiss, B. A., & Stapleton, L. (2014). An empirical evaluation of mediation effect analysis with manifest and latent variables using Markov Chain Monte Carlo and alternative estimation methods. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(2), 253-262. doi: https://doi.org/10.1080/10705511.2014.882688

Chen, J., Zhang, D., & Choi, J. (2015). Estimation of the latent mediated effect with ordinal data using the limited-information and Bayesian full-information

approaches. *Behavior research methods*, *47*(4), 1260-1273. doi: https://doi.org/10.3758/s13428-014-0526-3

Cheung, M. W. (2007). Comparison of approaches to constructing confidence intervals for mediating effects using structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(2), 227-246. doi: https://doi.org/10.1080/10705510709336745

Cheung, M. W. (2009). Comparison of methods for constructing confidence intervals of standardized indirect effects. *Behavior Research Methods*, *41*(2), 425-438. doi: https://doi.org/10.3758/brm.41.2.425

Cooper, W. H., & Richardson, A. J. (1986). Unfair comparisons. *Journal of Applied Psychology*, *71*(2), 179-184. doi: https://doi.org/10.1037//0021-9010.71.2.179

Cowles, M. K., & Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association, 91,* 883-904. doi: https://doi.org/10.1080/01621459.1996.10476956

Craig, C. C. (1936). On the frequency function of xy. *Annals of Mathematical Statistics*, *7*, 1-15. doi: https://doi.org/10.1214/aoms/1177732541

Daniels, M. J., & Kass, R. E. (1999). Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. *Journal of the American Statistical Association, 94*(448), 1254-1263. doi: https://doi.org/10.1080/01621459.1999.10473878

Daniels, M. J., Roy, J. A., Kim, C., Hogan, J. W., & Perri, M. G. (2012). Bayesian inference for the causal effect of mediation. *Biometrics*, *68*(4), 1028-1036. doi: https://doi.org/10.1111/j.1541-0420.2012.01781.x

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *1-17*. doi: https://doi.org/10.3389/fpsyg.2014.00781

Enders, C. K., Fairchild, A. J., & MacKinnon, D. P. (2013). A Bayesian approach for estimating mediation effects with missing data. *Multivariate Behavioral Research*, *48*(3), 340-369. doi: https://doi.org/10.1080/00273171.2013.784862

van Erp, S., Mulder, J., & Oberski, D. L. (2016). *Prior Sensitivity Analysis in Default Bayesian Structural Equation Modeling*. Manuscript submitted for publication.

Evans, M., & Jang, G. H. (2011). Weak informativity and the information in one prior relative to another. *Statistical Science, 26*(3), 423-439. doi: https://doi.org/10.1214/11-sts357

Falk, C. F., & Biesanz, J. C. (2015). Inference and interval estimation methods for indirect effects with latent variable models. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(1), 24-38. doi: https://doi.org/10.1080/10705511.2014.935266

Finch, J. F., West, S. G., & MacKinnon, D. P. (1997). Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models. *Structural Equation Modeling: A Multidisciplinary Journal*, *4*(2), 87-107. doi: https://doi.org/10.1080/10705519709540063

Fritz, M. S., Taylor, A. B., & MacKinnon, D. P. (2012). Explanation of two anomalous results in statistical mediation analysis. *Multivariate Behavioral Research, 47*, 61-87. doi: https://doi.org/10.1080/00273171.2012.640596

Gelman, A. (2003). A Bayesian formulation of exploratory data analysis and goodness-of-fit testing. *International Statistical Review, 71,* 369-382. doi: https://doi.org/10.1111/j.1751-5823.2003.tb00203.x

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis, 1,* 515-533. doi: https://doi.org/10.1214/06-ba117a

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis*. CRC press.

Gelman, A., & Meng, X. L. (1996). Model checking and model improvement. In *Markov chain Monte Carlo in practice* (pp. 189-201). Springer US.

Gelman, A., Meng, X. L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica, 6,* 733-807.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*, 457-472. doi: https://doi.org/10.1214/ss/1177011136

Gelman, A., & Shirley, K. (2011). Inference from simulations and monitoring convergence. In Brooks, S., Gelman, A., Jones, G. L., & Meng, X.-L. (Eds.), *Handbook of Markov Chain Monte Carlo* (pp. 116– 162). Boca Raton: Chapman and Hall/CRC.

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

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). Introducing Markov Chain Monte Carlo. In *Markov chain Monte Carlo in practice* (pp. 1-19). Springer US.

Heise, D. R. (1975). *Causal analysis*. John Wiley & Sons.

Hoyle, R. H. (2012). Introduction and Overview. In R. Hoyle (Ed.), *Handbook of structural equation modeling*. New York: Guilford Press.

Hoyle, R. H., & Kenny, D. A. (1999). Sample size, reliability, and tests of statistical mediation. *Statistical strategies for small sample research*, *1*, 195-222.

James, L. R., & Brett, J. M. (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology*, *69*(2), 307-321. doi: https://doi.org/10.1037//0021-9010.69.2.307

Jeffreys, H. (1961). *The theory of probability*. Oxford University Press.

Kaplan, D. (1988). The impact of specification error on the estimation, testing, and improvement of Structural Equation Models. *Multivariate Behavioral Research*, *23*(1), 69-86. doi: https://doi.org/10.1207/s15327906mbr2301_4

Kaplan, D., & Depaoli, S. (2012). Bayesian Structural Equation Modeling. In R. Hoyle (Ed.), *Handbook of structural equation modeling*. New York: Guilford Press.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90,* 773-795. doi: https://doi.org/10.1080/01621459.1995.10476572

Kass, R. E., & Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association, 90*(431), 928-934. doi: https://doi.org/10.1080/01621459.1995.10476592

Kuehl, K. S., Elliot, D. L., Goldberg, L., MacKinnon, D. P., Vila, B. J., Smith, J., ... & Sleigh, A. (2014). The safety and health improvement: enhancing law enforcement departments study: feasibility and findings. *Frontiers in public health*, *2*, 38-45. doi: https://doi.org/10.3389/fpubh.2014.00038

Ledgerwood, A., & Shrout, P. E. (2011). The trade-off between accuracy and precision in latent variable models of mediation processes. *Journal of personality and social psychology*, *101*(6), 1174-1188. doi: 10.1037/a0024776

Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach*. Chichester, England: Wiley.

Lee, S. Y., & Song, X. Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research*, *39*(4), 653-686. doi: https://doi.org/10.1207/s15327906mbr3904_4

Levy, R., & Choi, J. (2013). Bayesian Structural Equation Modeling. In G. R. Hancock, R. O. Mueller (Eds.), *Structural Equation Modeling: A Second Course* (2nd ed., pp. 563-623). Information Age Publishing.

Leonard, T., & Hsu, J. S. (1992). Bayesian inference for a covariance matrix. *The Annals of Statistics*, 1669-1696. doi: https://doi.org/10.1214/aos/1176348885

Link, W. A., & Eaton, M. J. (2012). On thinning of chains in MCMC. *Methods in Ecology and Evolution, 3* (1), 112-115. doi: https://doi.org/10.1111/j.2041-210x.2011.00131.x

Lomnicki, Z. A. (1967). On the distribution of products of random variables. *Journal of the Royal Statistical Society, 29*, 513-524.

Lunn, D.J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10, 325-337.

MacCallum, R. C., Edwards, M. C., & Cai, L. (2012). Hopes and cautions in implementing Bayesian structural equation modeling. *Psychological methods*, *17*(3), 340-345. doi: https://doi.org/10.1037/a0027131

MacKinnon, D. P. (2008). Introduction to statistical mediation analysis. Routledge.

MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007). Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods, 39*, 384-389. doi: https://doi.org/10.3758/bf03193007

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, *7*(1), 83-104. doi: https://doi.org/10.1037//1082-989x.7.1.83

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*(1), 99-128. doi: https://doi.org/10.1207/s15327906mbr3901_4

Manly, B. F. (1997). *Randomization, bootstrap and Monte Carlo methods in Biology*. Cornwall: T.J. International Ltd.

Meng, X. L. (1994). Posterior predictive p-values. *The Annals of Statistics*, 1142-1160. doi: https://doi.org/10.1214/aos/1176325622

Miočević, M., MacKinnon, D. P., & Levy, R. (2016). *Comparison of Bayesian and frequentist estimates of the mediated effect*. Manuscript under review.

Muthén, B., & Asparouhov, T. (2012a). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychological methods*,*17*(3), 313-335. doi: https://doi.org/10.1037/a0026802

Muthén, L.K., & Muthén, B.O. (1998-2015). Mplus User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén.

Neal, R. (2011). MCMC using Hamiltonian dynamics. In Brooks, S., Gelman, A., Jones, G. L., & Meng, X.-L. (Eds.), *Handbook of Markov Chain Monte Carlo* (pp. 116–162). Boca Raton: Chapman and Hall/CRC.

Nuijten, M. B., Wetzels, R., Matzke, D., Dolan, C. V., & Wagenmakers, E. J. (2015). A default Bayesian hypothesis test for mediation. *Behavior Research Methods*, *47*(1), 85-97. doi: https://doi.org/10.3758/s13428-014-0470-2

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Raftery, A. E. (1993). Bayesian model selection in structural equation modeling. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation modeling* (pp. 163- 180). Beverly Hills: Sage.

Raftery, A., Hoeting, J., Volinsky, C., Painter, I., & Yeung, K. Y. (2015). BMA: Bayesian Model Averaging. R package version 3.18.4. http://CRAN.R-project.org/package=BMA

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, *21*(2), 173-184. doi: https://doi.org/10.1177/01466216970212006

Rindskopf, D. (2012). Next steps in Bayesian structural equation models: Comments on, variations of, and extensions to Muthén and Asparouhov (2012). *Psychological Methods, 17*(3), 336–339. doi: https://doi.org/10.1037/a0027130

Roberts, G. O. (1996). Markov chain concepts related to sampling algorithms. In *Markov chain Monte Carlo in practice* (pp. 45-57). Springer US.

SAS Institute Inc. 2009. *SAS/STAT® 9.2 User's Guide, Second Edition*. Cary, NC: SAS Institute Inc.

Scheines, R., Hoijtink, H., & Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika, 64*, 37-52. doi: https://doi.org/10.1007/bf02294318

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: new procedures and recommendations. *Psychological methods*, *7*(4), 422-445. doi: https://doi.org/10.1037//1082-989x.7.4.422

Sinharay, S. (2004). Experiences with Markov chain Monte Carlo convergence assessment in two psychometric examples. *Journal of Educational and Behavioral Statistics, 29*(4), 461-488. doi: https://doi.org/10.3102/10769986029004461

Song, X.Y., & Lee, S.Y. (2012). Basic and Advanced Bayesian Structural Equation Modeling. John Wiley and Sons, Chichester, UK.

Song, X. Y., Xia, Y. M., Pan, J. H., & Lee, S. Y. (2011). Model comparison of Bayesian semiparametric and parametric structural equation models. *Structural Equation Modeling*, *18*(1), 55-72. doi: https://doi.org/10.1080/10705511.2011.532720

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583-639. doi: https://doi.org/10.1111/1467-9868.00353

Springer, M. D. & Thompson, W. E. (1966). The distribution of independent random variables. *SIAM Journal on Applied Mathematics, 14,* 511-526. doi: https://doi.org/10.1137/0114046

Stan Development Team. 2015. *Stan Modeling Language Users Guide and Reference Manual, Version 2.7.0.*

StataCorp. 2015. *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP.

Sturtz, S., Ligges, U., and Gelman, A. (2005). R2WinBUGS: A Package for Running WinBUGS from R. *Journal of Statistical Software, 12*(3), 1-16. doi: https://doi.org/10.18637/jss.v012.i03

Tofighi, D. & MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods, 43*, 692-700. doi: https://doi.org/10.3758/s13428-011-0076-x

Tofighi, D., West, S. G., MacKinnon, D. P. (2013). Multilevel mediation analysis: The effects of omitted variables in the 1–1–1 model. *British Journal of Mathematical and Statistical Psychology*, *66* (2), 290-307. doi: https://doi.org/10.1111/j.2044-8317.2012.02051.x

de la Torre, J., & Douglas, J. A. (2008). Model evaluation and multiple strategies in cognitive diagnosis: An analysis of fraction subtraction data. *Psychometrika*, *73*(4), 595-624. doi: https://doi.org/10.1007/s11336-008-9063-2

Valente, M. J., Gonzalez, O., Miočević, M., & MacKinnon, D. P. (2015). A Note on Testing Mediated Effects in Structural Equation Models Reconciling Past and Current Research on the Performance of the Test of Joint Significance. *Educational and Psychological Measurement*, 889-911. doi: https://doi.org/10.1177/0013164415618992

Wright, S. (1934). The method of path coefficients. *The Annals of Mathematical Statistics*, *5*(3), 161-215. doi: https://doi.org/10.1214/aoms/1177732676

Yang, R., & Berger, J. O. (1994). Estimation of a covariance matrix using the reference prior. *The Annals of Statistics*, 1195-1211. doi: https://doi.org/10.1214/aos/1176325625

Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological methods*, *14*(4), 301-322. doi: https://doi.org/10.1037/a0016972

Zhang, Z. (2013). Bayesian growth curve models with the generalized error distribution. *Journal of Applied Statistics*, *40*(8), 1779-1795. doi: https://doi.org/10.1080/02664763.2013.796348

Zhang, Z. (2014). WebBUGS: Conducting Bayesian Analysis online. *Journal of Statistical Software, 61*(7), 1-30. doi: https://doi.org/10.18637/jss.v061.i07

Zhang, Z., McArdle, J. J., Wang, L., & Hamagami, F. (2008). A SAS interface for Bayesian analysis with WinBUGS. *Structural Equation Modeling*, *15*(4), 705-728. doi: https://doi.org/10.1080/10705510802339106

APPENDIX A

COVARIANCE MATRICES OF MANIFEST INDICATORS AND LATENT VARIABLES FOR THE SINGLE MEDIATOR MODEL WITH LATENT VARIABLES

Covariance matrix of manifest indicators

	X1	X2	X3
X1	$\lambda^2_{X11}\varphi_{11}+\theta_{\delta_{11}}$		
X2	$\lambda_{X11}\varphi_{11}\lambda_{X21}$	$\lambda^2_{X21}\varphi_{11}+\theta_{\delta_{22}}$	
X3	$\lambda_{X11}\varphi_{11}\lambda_{X31}$	$\lambda_{X21}\varphi_{11}\lambda_{X31}$	$\lambda^2_{X31}\varphi_{11}+\theta_{\delta_{33}}$
Y1	$\lambda_{X11}\varphi_{11}\gamma_{11}\lambda_{Y11}$	$\lambda_{X21}\varphi_{11}\gamma_{11}\lambda_{Y11}$	$\lambda_{X31}\varphi_{11}\gamma_{11}\lambda_{Y11}$
Y2	$\lambda_{X11}\varphi_{11}\gamma_{11}\lambda_{Y21}$	$\lambda_{X21}\varphi_{11}\gamma_{11}\lambda_{Y21}$	$\lambda_{X31}\varphi_{11}\gamma_{11}\lambda_{Y21}$
Y3	$\lambda_{X11}\varphi_{11}\gamma_{11}\lambda_{Y31}$	$\lambda_{X21}\varphi_{11}\gamma_{11}\lambda_{Y31}$	$\lambda_{X31}\varphi_{11}\gamma_{11}\lambda_{Y31}$

	Y1	Y2	Y3
Y1	$\lambda^2_{Y11}(\gamma_{11}{}^2\varphi_{11}+\psi_{11})+\theta_{\varepsilon_{11}}$		
Y2	$\lambda_{Y11}(\gamma_{11}{}^2\varphi_{11}+\psi_{11})\lambda_{Y21}$	$\lambda^2_{Y21}(\gamma_{11}{}^2\varphi_{11}+\psi_{11})+\theta_{\varepsilon_{22}}$	
Y3	$\lambda_{Y11}(\gamma_{11}{}^2\varphi_{11}+\psi_{11})\lambda_{Y31}$	$\lambda_{Y21}(\gamma_{11}{}^2\varphi_{11}+\psi_{11})\lambda_{Y31}$	$\lambda^2_{Y31}(\gamma_{11}{}^2\varphi_{11}+\psi_{11})+\theta_{\varepsilon_{33}}$

	Y4
X1	$\lambda_{X11}\varphi_{11}(\gamma_{11}\beta_{21}+\gamma_{21})\lambda_{Y42}$
X2	$\lambda_{X21}\varphi_{11}(\gamma_{11}\beta_{21}+\gamma_{21})\lambda_{Y42}$
X3	$\lambda_{X31}\varphi_{11}(\gamma_{11}\beta_{21}+\gamma_{21})\lambda_{Y42}$
Y1	$\lambda_{Y11}\lambda_{Y42}(\gamma_{11}\beta_{21}+\gamma_{21})\gamma_{11}\varphi_{11}$
Y2	$\lambda_{Y21}\lambda_{Y42}(\gamma_{11}\beta_{21}+\gamma_{21})\gamma_{11}\varphi_{11}$
Y3	$\lambda_{Y31}\lambda_{Y42}(\gamma_{11}\beta_{21}+\gamma_{21})\gamma_{11}\varphi_{11}$
Y4	$\lambda^2_{Y42}(\gamma_{21}{}^2\varphi_{11}+2\gamma_{11}\beta_{21}\gamma_{21}\varphi_{11}+\beta_{21}{}^2(\gamma_{11}{}^2\varphi_{11}+\psi_{11})+\psi_{22})+\theta_{\varepsilon_{44}}$
Y5	$\lambda_{Y42}(\gamma_{21}{}^2\varphi_{11}+2\gamma_{11}\beta_{21}\gamma_{21}\varphi_{11}+\beta_{21}{}^2(\gamma_{11}{}^2\varphi_{11}+\psi_{11})+\psi_{22})\lambda_{Y52}$
Y6	$\lambda_{Y42}(\gamma_{21}{}^2\varphi_{11}+2\gamma_{11}\beta_{21}\gamma_{21}\varphi_{11}+\beta_{21}{}^2(\gamma_{11}{}^2\varphi_{11}+\psi_{11})+\psi_{22})\lambda_{Y62}$

	Y5
X1	$\lambda_{X11}\varphi_{11}(\gamma_{11}\beta_{21}+\gamma_{21})\lambda_{Y52}$

X2	$\lambda_{X21}\varphi_{11}(\gamma_{11}\beta_{21}+\gamma_{21})\lambda_{Y52}$
X3	$\lambda_{X31}\varphi_{11}(\gamma_{11}\beta_{21}+\gamma_{21})\lambda_{Y52}$
Y1	$\lambda_{Y11}\lambda_{Y52}(\gamma_{11}\beta_{21}+\gamma_{21})\gamma_{11}\varphi_{11}$
Y2	$\lambda_{Y21}\lambda_{Y52}(\gamma_{11}\beta_{21}+\gamma_{21})\gamma_{11}\varphi_{11}$
Y3	$\lambda_{Y31}\lambda_{Y52}(\gamma_{11}\beta_{21}+\gamma_{21})\gamma_{11}\varphi_{11}$
Y4	$\lambda_{Y42}(\gamma_{21}{}^2\varphi_{11}+2\gamma_{11}\beta_{21}\gamma_{21}\varphi_{11}+\beta_{21}{}^2(\gamma_{11}{}^2\varphi_{11}+\psi_{11})+\psi_{22})\lambda_{Y62}$
Y5	$\lambda_{Y52}^2(\gamma_{21}{}^2\varphi_{11}+2\gamma_{11}\beta_{21}\gamma_{21}\varphi_{11}+\beta_{21}{}^2(\gamma_{11}{}^2\varphi_{11}+\psi_{11})+\psi_{22})+\theta_{\varepsilon_{55}}$
Y6	$\lambda_{Y52}(\gamma_{21}{}^2\varphi_{11}+2\gamma_{11}\beta_{21}\gamma_{21}\varphi_{11}+\beta_{21}{}^2(\gamma_{11}{}^2\varphi_{11}+\psi_{11})+\psi_{22})\lambda_{Y62}$

	Y6
X1	$\lambda_{X11}\varphi_{11}(\gamma_{11}\beta_{21}+\gamma_{21})\lambda_{Y62}$
X2	$\lambda_{X21}\varphi_{11}(\gamma_{11}\beta_{21}+\gamma_{21})\lambda_{Y62}$
X3	$\lambda_{X31}\varphi_{11}(\gamma_{11}\beta_{21}+\gamma_{21})\lambda_{Y62}$
Y1	$\lambda_{Y11}\lambda_{Y62}(\gamma_{11}\beta_{21}+\gamma_{21})\gamma_{11}\varphi_{11}$
Y2	$\lambda_{Y21}\lambda_{Y62}(\gamma_{11}\beta_{21}+\gamma_{21})\gamma_{11}\varphi_{11}$
Y3	$\lambda_{Y31}\lambda_{Y62}(\gamma_{11}\beta_{21}+\gamma_{21})\gamma_{11}\varphi_{11}$
Y4	$\lambda_{Y42}(\gamma_{21}{}^2\varphi_{11}+2\gamma_{11}\beta_{21}\gamma_{21}\varphi_{11}+\beta_{21}{}^2(\gamma_{11}{}^2\varphi_{11}+\psi_{11})+\psi_{22})\lambda_{Y62}$
Y5	$\lambda_{Y52}(\gamma_{21}{}^2\varphi_{11}+2\gamma_{11}\beta_{21}\gamma_{21}\varphi_{11}+\beta_{21}{}^2(\gamma_{11}{}^2\varphi_{11}+\psi_{11})+\psi_{22})\lambda_{Y62}$
Y6	$\lambda_{Y62}^2(\gamma_{21}{}^2\varphi_{11}+2\gamma_{11}\beta_{21}\gamma_{21}\varphi_{11}+\beta_{21}{}^2(\gamma_{11}{}^2\varphi_{11}+\psi_{11})+\psi_{22})+\theta_{\varepsilon_{66}}$

Covariance matrix of latent variables

	ξ	η_1	η_2
ξ	φ_{11}		
η_1	$\gamma_{11}\varphi_{11}$	$\gamma_{11}{}^2\varphi_{11}+\psi_{11}$	
η_2	$(\gamma_{11}\beta_{21}+\gamma_{21})\varphi_{11}$	$(\gamma_{11}\beta_{21}+\gamma_{21})\gamma_{11}\varphi_{11}$	$\gamma_{21}{}^2\varphi_{11}+2\gamma_{11}\beta_{21}\gamma_{21}\varphi_{11}+\beta_{21}{}^2(\gamma_{11}{}^2\varphi_{11}+\psi_{11})+\psi_{22}$

Covariance algebra for the covariance matrix of manifest indicators

$$Cov(X_1, X_1) = Cov(\lambda_{X11}\xi + \delta_1, \lambda_{X11}\xi + \delta_1) = \lambda^2_{X11}\varphi_{11} + \theta_{\delta_{11}}$$

$$Cov(X_1, X_2) = Cov(\lambda_{X11}\xi + \delta_1, \lambda_{X21}\xi + \delta_2) = \lambda_{X11}\varphi_{11}\lambda_{X21}$$

$$Cov(X_1, X_3) = Cov(\lambda_{X11}\xi + \delta_1, \lambda_{X31}\xi + \delta_3) = \lambda_{X11}\varphi_{11}\lambda_{X31}$$

$$Cov(X_1, Y_1) = Cov(\lambda_{X11}\xi + \delta_1, \lambda_{Y11}\eta_1 + \varepsilon_1) = \lambda_{X11}Cov(\xi, \eta_1)\lambda_{Y11} = \lambda_{X11}(\gamma_{11}\varphi_{11})\lambda_{Y11}$$

$$Cov(X_1, Y_2) = Cov(\lambda_{X11}\xi + \delta_1, \lambda_{Y21}\eta_1 + \varepsilon_2) = \lambda_{X11}Cov(\xi, \eta_1)\lambda_{Y21} = \lambda_{X11}(a\varphi_{11})\lambda_{Y21}$$

$$Cov(X_1, Y_3) = Cov(\lambda_{X11}\xi + \delta_1, \lambda_{Y31}\eta_1 + \varepsilon_3) = \lambda_{X11}Cov(\xi, \eta_1)\lambda_{Y31} = \lambda_{X11}(\gamma_{11}\varphi_{11})\lambda_{Y31}$$

$$Cov(X_1, Y_4) = Cov(\lambda_{X11}\xi + \delta_1, \lambda_{Y42}\eta_2 + \varepsilon_4) = \lambda_{X11}Cov(\xi, \eta_2)\lambda_{Y42} = \lambda_{X11}(\gamma_{11}\beta_{21} + \gamma_{21})\varphi_{11}\lambda_{Y42}$$

$$Cov(X_1, Y_5) = Cov(\lambda_{X11}\xi + \delta_1, \lambda_{Y52}\eta_2 + \varepsilon_5) = \lambda_{X11}Cov(\xi, \eta_2)\lambda_{Y52}$$
$$= \lambda_{X11}(\gamma_{11}\beta_{21} + \gamma_{21})\varphi_{11}\lambda_{Y52}$$

$$Cov(X_1, Y_6) = Cov(\lambda_{X11}\xi + \delta_1, \lambda_{Y62}\eta_2 + \varepsilon_6) = \lambda_{X11}Cov(\xi, \eta_2)\lambda_{Y62}$$
$$= \lambda_{X11}(\gamma_{11}\beta_{21} + \gamma_{21})\varphi_{11}\lambda_{Y62}$$

Note: covariance algebra for indicators X_2 and X_3 is omitted, as it is identical to the covariance algebra for X_1, but with different subscripts

$$Cov(Y_1, Y_1) = Cov(\lambda_{Y11}\eta_1 + \varepsilon_1, \lambda_{Y11}\eta_1 + \varepsilon_1) = \lambda^2{}_{Y11}(\gamma_{11}{}^2\varphi_{11} + \psi_{11}) + \theta_{\varepsilon_{11}}$$

$$Cov(Y_1, Y_2) = Cov(\lambda_{Y11}\eta_1 + \varepsilon_1, \lambda_{Y21}\eta_1 + \varepsilon_2) = \lambda_{Y11}(\gamma_{11}{}^2\varphi_{11} + \psi_{11})\lambda_{Y21}$$

$$Cov(Y_1, Y_3) = Cov(\lambda_{Y11}\eta_1 + \varepsilon_1, \lambda_{Y31}\eta_1 + \varepsilon_3) = \lambda_{Y11}(\gamma_{11}{}^2\varphi_{11} + \psi_{11})\lambda_{Y31}$$

$$Cov(Y_1, Y_4) = Cov(\lambda_{Y11}\eta_1 + \varepsilon_1, \lambda_{Y42}\eta_2 + \varepsilon_2) = \lambda_{Y11}(\gamma_{11}\beta_{21} + \gamma_{21})\gamma_{11}\varphi_{11}\lambda_{Y42}$$

$$Cov(Y_1, Y_5) = Cov(\lambda_{Y11}\eta_1 + \varepsilon_1, \lambda_{Y52}\eta_2 + \varepsilon_5) = \lambda_{Y11}(\gamma_{11}\beta_{21} + \gamma_{21})\gamma_{11}\varphi_{11}\lambda_{Y52}$$

$$Cov(Y_1, Y_6) = Cov(\lambda_{Y11}\eta_1 + \varepsilon_1, \lambda_{Y62}\eta_2 + \varepsilon_6) = \lambda_{Y11}(\gamma_{11}\beta_{21} + \gamma_{21})\gamma_{11}\varphi_{11}\lambda_{Y62}$$

Note: covariance algebra for indicators Y_2 and Y_3 is omitted, as it is identical to the covariance algebra for Y_1, but with different subscripts

$$Cov(Y_4, Y_4) = Cov(\lambda_{Y42}\eta_2 + \varepsilon_4, \lambda_{Y42}\eta_2 + \varepsilon_4) = \lambda^2{}_{Y42} Var(\eta_2) + \theta_{\varepsilon_4} =$$
$$\lambda^2{}_{Y42}(\gamma_{21}{}^2\varphi_{11} + 2\gamma_{11}\beta_{21}\gamma_{21}\varphi_{11} + \beta_{21}{}^2(\gamma_{11}{}^2\varphi_{11} + \psi_{11}) + \psi_{22}) + \theta_{\varepsilon_4}$$

$$Cov(Y_4, Y_5) = Cov(\lambda_{Y42}\eta_2 + \varepsilon_4, \lambda_{Y52}\eta_2 + \varepsilon_5) = \lambda_{Y42} Var(\eta_2)\lambda_{Y52} =$$
$$\lambda_{Y42}(\gamma_{21}{}^2\varphi_{11} + 2\gamma_{11}\beta_{21}\gamma_{21}\varphi_{11} + \beta_{21}{}^2(\gamma_{11}{}^2\varphi_{11} + \psi_{11}) + \psi_{22})\lambda_{Y52}$$

$$Cov(Y_4, Y_5) = Cov(\lambda_{Y42}\eta_2 + \varepsilon_4, \lambda_{Y62}\eta_2 + \varepsilon_6) = \lambda_{Y42} Var(\eta_2)\lambda_{Y62} =$$
$$\lambda_{Y42}(\gamma_{21}{}^2\varphi_{11} + 2\gamma_{11}\beta_{21}\gamma_{21}\varphi_{11} + \beta_{21}{}^2(\gamma_{11}{}^2\varphi_{11} + \psi_{11}) + \psi_{22})\lambda_{Y62}$$

Note: covariance algebra for indicators Y_5 and Y_6 is omitted, as it is identical to the covariance algebra for Y_4, but with different subscripts

Covariance algebra for the covariance matrix of latent variables

$$Var(\xi) = \varphi_{11}$$

$$Cov(\xi, \eta_1) = Cov(\xi, \gamma_{11}\xi + \zeta_1) = \gamma_{11}\varphi_{11}$$

$$Cov(\xi, \eta_1) = Cov(\xi, \gamma_{21}\xi + \beta_{21}\eta_1 + \zeta_2) = \gamma_{21}\varphi_{11} + \beta_{21} Cov(\xi, \eta_1)$$
$$= \gamma_{21}\varphi_{11} + \gamma_{11}\beta_{21}\varphi_{11} = (\gamma_{11}\beta_{21} + \gamma_{21})\varphi_{11}$$

$$Var(\eta_1) = Cov(\gamma_{11}\xi + \zeta_1, \gamma_{11}\xi + \zeta_1) = \gamma_{11}{}^2\varphi_{11} + \psi_{11}$$

$$Cov(\eta_1, \eta_2) = Cov(\gamma_{11}\xi + \zeta_1, \gamma_{21}\xi + \beta_{21}\eta_1 + \zeta_2) = \gamma_{11}\gamma_{21}\varphi_{11} + \gamma_{11}\beta_{21} Cov(\xi, \eta_1)$$
$$= \gamma_{11}\gamma_{21}\varphi_{11} + \gamma_{11}\beta_{21}\gamma_{11}\varphi_{11} = (\gamma_{11}\beta_{21} + \gamma_{21})\gamma_{11}\varphi_{11}$$

$$Var(\eta_2) = Cov(\gamma_{21}\xi + \beta_{21}\eta_1 + \zeta_2, \gamma_{21}\xi + \beta_{21}\eta_2 + \zeta_2)$$
$$= \gamma_{21}{}^2\varphi_{11} + 2\gamma_{11}\beta_{21}\gamma_{21}\varphi_{11} + \beta_{21}{}^2(\gamma_{11}{}^2\varphi_{11} + \psi_{11}) + \psi_{22}$$

APPENDIX B

COMPUTING AND MANUPULATING INDICATOR RELIABILITY IN THE MONTE CARLO STUDY

In order to determine the appropriate model parameters for generating a population with latent X, M, and Y, with three indicators per latent variable and indicators of the desired reliability level ρ, it was necessary to derive the variances of the indicators and latent variables, and to express the residual variances of indicators as functions of the remaining model parameters. Appendix A contains details of the derivation using covariance algebra and Wright's (1934) path tracing rules expanded to describe tracing rules for unstandardized variables (Heise, 1975). This appendix contains the calculations for manipulating reliabilities of manifest indicators for a population where all latent variables in the model, i.e., ξ, η_1, and η_2 all have variances of 1, and where the structural coefficients γ_{11}, β_{21}, and γ_{21} are selected by the researcher. The desired level of reliability for a given indicator was obtained by adjusting the true values of residual variances ψ_{11} and ψ_{22}.

$$Var(\xi) = \varphi_{11}$$

$$Var(\eta_1) = \gamma_{11}{}^2\varphi_{11} + \psi_{11}$$

$$Var(\eta_2) = \gamma_{21}{}^2\varphi_{11} + 2\gamma_{11}\beta_{21}\gamma_{21}\varphi_{11} + \beta_{21}{}^2(\gamma_{11}{}^2\varphi_{11} + \psi_{11}) + \psi_{22}$$

$$Var(X_1) = \lambda_{X11}^2\varphi_{11} + \theta_{\delta_{11}}$$

$$Var(X_2) = \lambda_{X21}^2\varphi_{11} + \theta_{\delta_{22}}$$

$$Var(X_3) = \lambda_{X31}^2\varphi_{11} + \theta_{\delta_{33}}$$

$$Var(Y_1) = \lambda_{Y11}^2 Var(\eta_1) + \theta_{\varepsilon_{11}}$$

$$Var(Y_2) = \lambda_{Y21}^2 Var(\eta_1) + \theta_{\varepsilon_{22}}$$

$$Var(Y_3) = \lambda_{Y31}^2 Var(\eta_1) + \theta_{\varepsilon_{33}}$$

$$Var(Y_4) = \lambda_{Y42}^2 Var(\eta_2) + \theta_{\varepsilon_{44}}$$

$$Var(Y_5) = \lambda_{Y52}^2 Var(\eta_2) + \theta_{\varepsilon_{55}}$$

$$Var(Y_6) = \lambda_{Y62}^2 Var(\eta_2) + \theta_{\varepsilon_{66}}$$

The formula for reliability of the indicator X_1 then becomes:

$$\rho_{X_1} = \frac{\lambda_{X11}^2 \varphi_{11}}{\lambda_{X11}^2 \varphi_{11} + \theta_{\delta_{11}}}$$

Note, the formula for the reliability of indicators X_2 and X_3 is the analogous, so for the case of equally reliable indicator variables only the computation for $\theta_{\delta_{11}}$ from already chosen λ_{X11}, φ_{11}, and ρ_{X_1} is presented. For $\lambda_{X11} = 1$, $\varphi_{11} = 1$, and $\rho_{X_1} = .7$ the calculation of $\theta_{\delta_{11}}$ becomes:

$$.7 = \frac{1 \cdot 1}{1 \cdot 1 + \theta_{\delta_{11}}}$$

$$1 + \theta_{\delta_{11}} = \frac{1}{.7}$$

$$\theta_{\delta_{11}} = \frac{1}{.7} - 1 = .4286$$

The formula for reliability of the indicators Y_1 and Y_4 then become:

$$\rho_{Y_1} = \frac{\lambda_{Y11}^2 Var(\eta_1)}{\lambda_{Y11}^2 Var(\eta_1) + \theta_{\varepsilon_{11}}}$$

$$\rho_{Y_4} = \frac{\lambda_{Y42}^2 Var(\eta_2)}{\lambda_{Y42}^2 Var(\eta_2) + \theta_{\varepsilon_{44}}}$$

Note, the formula for the reliability of all of the manifest indicators in the model is the same, so the residual variances of all indicators when reliability is 0.7 is 0.4286. The

parameters that changes in order to maintain the variance of 1 for latent variables LM, and LY are the residual variances ψ_{11} and ψ_{22}. The formulae for computing their values for the population-generating model are:

$\psi_{11} = 1 - \gamma_{11}^{\ 2}$ for $\varphi_{11} = 1$

$\psi_{22} = 1 - (\gamma_{21}^{\ 2} + 2\gamma_{11}\beta_{21}\gamma_{21} + \beta_{21}^{\ 2})$ for $\varphi_{11} = 1$ and $Var(\eta_1) = 1$

APPENDIX C

SAMPLE CODE FROM THE MONTE CARLO STUDY

This code was used in the Monte Carlo study for this project to generate data, separate
data into samples for iterations, to obtain parameter estimates using ML and the
distribution of the product, to obtain parameter summaries in the Bayesian framework
with diffuse conjugate priors, and to obtain parameter summaries in the Bayesian
framework with accurate informative priors for N_{prior} = 100.
The example code comes from Combination 1 of the Monte Carlo study.
The structure of the code is based on syntax by Dr. Roy Levy.

Data Generation

```
# Call the packages

library(MASS)
library(lavaan)

# Define the covariance matrix of the latent variables

elphi <-c(1,a,(a*b+cp),a,1,(a*cp+b),a*b+cp,(a*cp+b),1)
phi <- matrix(elphi, nrow = M, ncol = M)

# Define the loading matrix

load<-c(loading, loading, loading, 0,0,0,0,0,0,
        0,0,0,loading,loading,loading,0,0,0,
        0,0,0,0,0,0,loading,loading,loading)
lambda <- matrix(load, nrow=J,ncol=M)

# Define the structural coefficients matrix

cols <- c(0,a,cp,0,0,b,0,0,0)

beta <- matrix(cols, nrow=M, ncol=M)

# Define the covariance matrix of the errors for the indicators

miresvar <- .4286
indresvar<-c(miresvar, 0, 0, 0,0,0,0,0,0,
        0,miresvar,0,0,0,0,0,0,0,
        0,0,miresvar,0,0,0,0,0,0,
        0,0,0,miresvar,0,0,0,0,0,
        0,0,0,0,miresvar,0,0,0,0,
        0,0,0,0,0,miresvar,0,0,0,
        0,0,0,0,0,0,miresvar,0,0,
        0,0,0,0,0,0,0,miresvar,0,
        0,0,0,0,0,0,0,0,miresvar)
psi.x <- matrix(indresvar, nrow=J,ncol=J)

var.ind <- lambda%*%phi%*%t(lambda)+psi.x

# Define the model implied covariance structure of the observables

cov.lat.ind<-lambda%*%phi
```

```
# Define the model implied covariance structure of the ALL the
# variables

  # Initiate matrix
  covariance.latent.and.observables <- matrix(NA, nrow=M+J, ncol=M+J)

  # Fill in cov matrix of latents
  covariance.latent.and.observables[1:M, 1:M] <- phi

  # Fill in cov matrix of latents with observables
  covariance.latent.and.observables[1:M, (M+1):(M+J)] <- t(cov.lat.ind)
  covariance.latent.and.observables[(M+1):(M+J), 1:M] <- cov.lat.ind

  # Fill in cov matrix of observables
  covariance.latent.and.observables[(M+1):(M+J), (M+1):(M+J)] <-
        var.ind

# Loop over replications and generate the data

which.rep=0
which.rep=which.rep+1

for(which.rep in 1:n.reps){

# Generate multivariate normal of right size

generated.data.raw <- mvrnorm(n=n, mu=rep(0, M+J),
Sigma=covariance.latent.and.observables)

# Store the generated data
generated.data <- generated.data.raw

# Label the columns
  column.names <- c(
    paste("Ksi", seq(1:M), sep=""),
    paste("x", seq(1:J), sep="")
  )
  colnames(generated.data) <- column.names

data.file.name <- paste("gen.data.", which.rep, ".csv", sep="")

  write.csv(
    x=generated.data,
    file=data.file.name,
    row.names=FALSE,
    col.names=TRUE
  )

} # closes loop over replications
```

Data Separation

```
# Code to read in and separate the data
# Should be called by another file
# where the replication is contained in "which.rep"
```

```
# and the data folder has already been defined

# Name the data file based on the replication
data.file.name <- paste("gen.data.", which.rep, ".csv", sep="")

# Read in all the variables
all.variables <- read.csv(paste(condition.data.folder, data.file.name,
sep=""), header=TRUE)
```

Analysis using ML and distribution of the product

```
# Call the 'lavaan' library

library(lavaan)

# Loop over replications

which.rep=0
which.rep=which.rep+1

sum.stats.all = data.frame(which.rep=rep(0, n.reps), ab.ml=rep(0,
n.reps), ab.ml.se=rep(0, n.reps),
                           bias.ml=rep(0, n.reps), rel.bias.ml=rep(0,
n.reps), ab.low.ml=rep(0, n.reps),
                           ab.up.ml=rep(0, n.reps), pow.ml=rep(0,
n.reps), coverage.ml=rep(0, n.reps),
                           typei.ml=rep(0, n.reps), imb.ml=rep(0,
n.reps), int.width.ml=rep(0, n.reps))


for(which.rep in 1:n.reps){

# Read in the data

if(1==1){
    file.name <- "Read in and Separate Data.R"
    source(paste(drive.letter, file.name, sep=""))
  }

# Data

raw.data = read.csv(data.file.name)

model <- '
            X =~ x1 + x2 + x3
            M =~ x4 + x5 + x6
            Y =~ x7 + x8 + x9

# direct effect
             Y ~ c*X
           # mediator
             M ~ a*X
             Y ~ b*M
           # indirect effect (a*b)
```

```
            ab := a*b
          # total effect
            total := c + (a*b)
        '
fit.ml <- sem(model, data = raw.data)
summary(fit.ml, fit.measures = TRUE)

ab.ml <- parameterEstimates(fit.ml)$est[25]
ab.ml.se <- parameterEstimates(fit.ml)$se[25]

true.ab<-t.a*t.b

# bias
bias.ml<-ab.ml-true.ab

# relative bias
rel.bias.ml<-bias.ml/true.ab


# Distribution of the product confidence limits

library(RMediation)

a.ml <- parameterEstimates(fit.ml)$est[11]
a.ml.se <- parameterEstimates(fit.ml)$se[11]
b.ml <- parameterEstimates(fit.ml)$est[12]
b.ml.se <- parameterEstimates(fit.ml)$se[12]

dop.obj<-medci(a.ml, b.ml, a.ml.se, b.ml.se, rho = 0, alpha = 0.05,
type = "dop",
      plot=FALSE, plotCI=FALSE)


ab.low.ml <- dop.obj$`97.5% CI`[[1]]
ab.up.ml <- dop.obj$`97.5% CI`[[2]]


# interval outcomes
pow.ml <- NA
if (ab.low.ml > 0 & ab.up.ml > 0) {
  pow.ml <- 1
} else {
  pow.ml <- 0
}
coverage.ml <- NA
if (true.ab > ab.low.ml & true.ab < ab.up.ml) {
  coverage.ml <- 1
} else {
  coverage.ml<-0
}
typei.ml <- NA
imb.ml <- NA
imb.r.ml <- NA
imb.l.ml <- NA
  if (true.ab > ab.up.ml) {
```

```
    imb.r.ml <- 1
  } else {
    imb.r.ml <- 0
  }
if (true.ab < ab.low.ml) {
  imb.l.ml <- 1
} else {
  imb.l.ml <- 0
}
imb.ml <- imb.r.ml-imb.l.ml
int.width.ml <- ab.up.ml - ab.low.ml

sum.stats <- cbind(which.rep, ab.ml, ab.ml.se,
                   bias.ml, rel.bias.ml, ab.low.ml,
                   ab.up.ml, pow.ml, coverage.ml,
                   typei.ml, imb.ml, int.width.ml)

sum.stats.name <- paste("sum.stats",which.rep,sep="")
assign(sum.stats.name, sum.stats)
names(sum.stats) <- sum.stats.name

sum.stats.all[which.rep, ] <- sum.stats

} # closes loop over replications

# Write out the summary statistics

write.csv(
  x=sum.stats.all,
  file="Summary Statistics ML Comb1.csv"
)
```

Analysis using Bayesian methods with diffuse conjugate priors

```
# Call the 'R2WinBUGS' library
library(R2WinBUGS)
# Loop over replications

which.rep=0
which.rep=which.rep+1

sum.stats.all = data.frame(which.rep=rep(0, n.reps),
ab.mean=rep(0,n.reps), ab.med=rep(0,n.reps),
                           ab.low.hpd=rep(0,n.reps),
ab.up.hpd=rep(0,n.reps), try.mean=rep(0,n.reps),
                           bias.mean=rep(0,n.reps),
bias.med=rep(0,n.reps), ab.sd=rep(0,n.reps),
                           rel.bias.mean=rep(0,n.reps),
rel.bias.med=rep(0,n.reps),
                           pow=rep(0,n.reps), typei=rep(0,n.reps),
coverage=rep(0,n.reps),
                           imb=rep(0,n.reps), int.width=rep(0,n.reps))

for(which.rep in 1:n.reps){
```

```
# Read in the data

  if(1==1){
    file.name <- "Read in and Separate Data.R"
    source(paste(drive.letter, file.name, sep=""))
  }
# Define the model
# by creating a string that is the BUGS code

modelstring <- as.character("

model{

# Prior distributions

# Measurement model

lam.x2 ~ dnorm(0, tau.x2);
lam.x3 ~ dnorm(0, tau.x3);

lam.m2 ~ dnorm(0, tau.m2);
lam.m3 ~ dnorm(0, tau.m3);

lam.y2 ~ dnorm(0, tau.y2);
lam.y3 ~ dnorm(0, tau.y3);

tau.x1 ~ dgamma(.5, .5);
tau.x2 ~ dgamma(.5, .5);
tau.x3 ~ dgamma(.5, .5);

sigma.x1 <-1/tau.x1;
sigma.x2 <-1/tau.x2;
sigma.x3 <-1/tau.x3;

tau.m1 ~ dgamma(.5, .5);
tau.m2 ~ dgamma(.5, .5);
tau.m3 ~ dgamma(.5, .5);

sigma.m1 <-1/tau.m1;
sigma.m2 <-1/tau.m2;
sigma.m3 <-1/tau.m3;

tau.y1 ~ dgamma(.5, .5);
tau.y2 ~ dgamma(.5, .5);
tau.y3 ~ dgamma(.5, .5);

sigma.y1 <-1/tau.y1;
sigma.y2 <-1/tau.y2;
sigma.y3 <-1/tau.y3;

# Structural model
tau.latx ~ dgamma(.5, .5);
tau.res.latm ~ dgamma(.5, .5);
tau.res.laty ~ dgamma(.5, .5);
```

```
a ~ dnorm(0, tau.res.latm);
b ~ dnorm(0, tau.res.laty);
cp ~ dnorm(0, tau.res.laty);

ab <-a*b

# Conditional probability of the data
# A regression model

for(i in 1:n){
x1.mean[i] <- 1*Ksi1[i];
x2.mean[i] <- lam.x2*Ksi1[i];
x3.mean[i] <- lam.x3*Ksi1[i];

x4.mean[i] <- 1*Ksi2[i];
x5.mean[i] <- lam.m2*Ksi2[i];
x6.mean[i] <- lam.m3*Ksi2[i];

x7.mean[i] <- 1*Ksi3[i];
x8.mean[i] <- lam.y2*Ksi3[i];
x9.mean[i] <- lam.y3*Ksi3[i];

Ksi1[i] ~ dnorm(0,tau.latx)

x1[i] ~ dnorm(x1.mean[i], tau.x1);
x2[i] ~ dnorm(x2.mean[i], tau.x2);
x3[i] ~ dnorm(x3.mean[i], tau.x3);

x4[i] ~ dnorm(x4.mean[i], tau.m1);
x5[i] ~ dnorm(x5.mean[i], tau.m2);
x6[i] ~ dnorm(x6.mean[i], tau.m3);

x7[i] ~ dnorm(x7.mean[i], tau.y1);
x8[i] ~ dnorm(x8.mean[i], tau.y2);
x9[i] ~ dnorm(x9.mean[i], tau.y3);

Ksi2.mean[i] <- a*Ksi1[i];
Ksi3.mean[i] <- cp*Ksi1[i] + b*Ksi2[i];

Ksi2[i] ~ dnorm(Ksi2.mean[i],tau.res.latm)
Ksi3[i] ~ dnorm(Ksi3.mean[i],tau.res.laty)
       }
}

")

# Write out the BUGS code to a file

BUGS.code.file.name <- "Diffuse.txt"
write(modelstring, BUGS.code.file.name)
write(modelstring, "temp.bug")

# Define data to give to BUGS

raw.data = read.csv(data.file.name)
```

```
n <- nrow(raw.data)
x1 = raw.data$x1
x2 = raw.data$x2
x3 = raw.data$x3
x4 = raw.data$x4
x5 = raw.data$x5
x6 = raw.data$x6
x7 = raw.data$x7
x8 = raw.data$x8
x9 = raw.data$x9

data <- list("n", "x1", "x2", "x3", "x4", "x5", "x6", "x7", "x8", "x9")

# Define parameters to monitor in BUGS

parameters <- c("lam.x2", "lam.x3", "tau.x1", "tau.x2", "tau.x3",
                "lam.m2", "lam.m3", "tau.m1", "tau.m2", "tau.m3",
                "lam.y2", "lam.y3", "tau.y1", "tau.y2", "tau.y3",
                "tau.latx", "tau.res.latm", "tau.res.laty", "a", "b",
"cp", "ab")

# Define initial values to give to BUGS for each of 3 chains

lam.x2.inits.1 = 0
lam.x3.inits.1 = 0.3
tau.x1.inits.1 = 2
tau.x2.inits.1 = 1
tau.x3.inits.1 = .5

lam.m2.inits.1 = 0
lam.m3.inits.1 = 0.3
tau.m1.inits.1 = 2
tau.m2.inits.1 = 1
tau.m3.inits.1 = .5

lam.y2.inits.1 = 0
lam.y3.inits.1 = 0.3
tau.y1.inits.1 = 2
tau.y2.inits.1 = 1
tau.y3.inits.1 = .5

tau.latx.inits.1= .5
tau.res.latm.inits.1= .5
tau.res.laty.inits.1= .5
a.inits.1= .3
b.inits.1= .4
cp.inits.1= .7

inits1 <- list(lam.x2=lam.x2.inits.1, lam.x3=lam.x3.inits.1,
               tau.x1=tau.x1.inits.1, tau.x2=tau.x2.inits.1,
tau.x3=tau.x3.inits.1,
               lam.m2=lam.m2.inits.1, lam.m3=lam.m3.inits.1,
               tau.m1=tau.m1.inits.1, tau.m2=tau.m2.inits.1,
tau.m3=tau.m3.inits.1,
               lam.y2=lam.y2.inits.1, lam.y3=lam.y3.inits.1,
```

```
              tau.y1=tau.y1.inits.1, tau.y2=tau.y2.inits.1,
tau.y3=tau.y3.inits.1,
              tau.latx=tau.latx.inits.1,
tau.res.latm=tau.res.latm.inits.1,
              tau.res.laty=tau.res.laty.inits.1, a=a.inits.1,
b=b.inits.1, cp=cp.inits.1)

lam.x2.inits.2 = 0.3
lam.x3.inits.2 = 0.1
tau.x1.inits.2 = .5
tau.x2.inits.2 = 2
tau.x3.inits.2 = 1

lam.m2.inits.2 = 0.3
lam.m3.inits.2 = 0
tau.m1.inits.2 = 1
tau.m2.inits.2 = .5
tau.m3.inits.2 = 2

lam.y2.inits.2 = 0.3
lam.y3.inits.2 = 0
tau.y1.inits.2 = 1
tau.y2.inits.2 = .5
tau.y3.inits.2 = 2

tau.latx.inits.2= 1
tau.res.latm.inits.2= 1
tau.res.laty.inits.2= 1
a.inits.2= .7
b.inits.2= .3
cp.inits.2= .4

inits2 <- list(lam.x2=lam.x2.inits.2, lam.x3=lam.x3.inits.2,
              tau.x1=tau.x1.inits.2, tau.x2=tau.x2.inits.2,
tau.x3=tau.x3.inits.2,
              lam.m2=lam.m2.inits.2, lam.m3=lam.m3.inits.2,
              tau.m1=tau.m1.inits.2, tau.m2=tau.m2.inits.2,
tau.m3=tau.m3.inits.2,
              lam.y2=lam.y2.inits.2, lam.y3=lam.y3.inits.2,
              tau.y1=tau.y1.inits.2, tau.y2=tau.y2.inits.2,
tau.y3=tau.y3.inits.2,
              tau.latx=tau.latx.inits.2,
tau.res.latm=tau.res.latm.inits.2,
              tau.res.laty=tau.res.laty.inits.2, a=a.inits.2,
b=b.inits.2, cp=cp.inits.2)

lam.x2.inits.3 = 0.01
lam.x3.inits.3 = 0
tau.x1.inits.3 = 1
tau.x2.inits.3 = .5
tau.x3.inits.3 = 2

lam.m2.inits.3 = 0.3
lam.m3.inits.3 = 0.7
tau.m1.inits.3 = 2.4
```

```
tau.m2.inits.3 = 1.7
tau.m3.inits.3 = .2

lam.y2.inits.3 = 0.3
lam.y3.inits.3 = 3
tau.y1.inits.3 = 2.5
tau.y2.inits.3 = 1.7
tau.y3.inits.3 = .85

tau.latx.inits.3= .15
tau.res.latm.inits.3= .75
tau.res.laty.inits.3= 3.5
a.inits.3= 2.3
b.inits.3= 1.4
cp.inits.3= .2


inits3 <- list(lam.x2=lam.x2.inits.3, lam.x3=lam.x3.inits.3,
               tau.x1=tau.x1.inits.3, tau.x2=tau.x2.inits.3,
tau.x3=tau.x3.inits.3,
               lam.m2=lam.m2.inits.3, lam.m3=lam.m3.inits.3,
               tau.m1=tau.m1.inits.3, tau.m2=tau.m2.inits.3,
tau.m3=tau.m3.inits.3,
               lam.y2=lam.y2.inits.3, lam.y3=lam.y3.inits.3,
               tau.y1=tau.y1.inits.3, tau.y2=tau.y2.inits.3,
tau.y3=tau.y3.inits.3,
               tau.latx=tau.latx.inits.3,
tau.res.latm=tau.res.latm.inits.3,
               tau.res.laty=tau.res.laty.inits.3, a=a.inits.3,
b=b.inits.3, cp=cp.inits.3)

inits <- list(inits1, inits2, inits3)

# Choose
# the number of chains
# the number of iterations to burn-in,
# the number of iterations to thin by,
# the total number of iterations

n.chains = 3
n.burnin = 1500
n.thin = 1
n.iters.total.per.chain = 5000

# Call WinBUGS to run the model, returning a BUGS object

model.in.winbugs <- bugs(
  data=data,
  inits=inits,
  parameters.to.save=parameters,
  model.file=BUGS.code.file.name,
  n.chains=n.chains,
  n.iter=n.iters.total.per.chain,
        n.burnin=n.burnin,
        n.thin=n.thin,
```

```
        debug=FALSE,
        codaPkg=TRUE,
  working.directory=getwd(),
  bugs.directory = 'C:\\Program Files\\WinBUGS14',
        DIC=FALSE,
)

# This code reads in the draws from BUGS

library(coda)

# Define the R object with the names of the coda files
coda.file.names <- model.in.winbugs
# Read the files into coda
draws.from.bugs <- read.bugs(coda.file.names)
# Convert the draws to a matrix
draws.from.bugs.as.matrix <- as.matrix(draws.from.bugs)

# This code summarizes the draws
# Combine chains for summaries

coda.options(combine.stats=TRUE, combine.plots=TRUE)

# Extract the summary statistics
#   Usual
#   Percentiles
#   HPD

export.draws <- do.call(rbind.data.frame, draws.from.bugs)
export.draws.name <- paste("diffuse_conjugate_draws.", which.rep,
".csv", sep="")

write.csv(x=export.draws,
  file=export.draws.name,
  row.names=FALSE,
  col.names=TRUE)
summary.stats <- summary(draws.from.bugs)

stats<- as.data.frame(summary.stats$statistics)
ab.mean <- stats$Mean[2] # this is the post mean of ab

#summary.stats$statistics
quant <-as.data.frame(summary.stats$quantiles)

ab.med <- quant[2,3] #post median of ab

true.ab<-a*b
# bias
bias.mean<-ab.mean-true.ab
bias.med<-ab.med-true.ab
# relative bias
rel.bias.mean<-bias.mean/true.ab
rel.bias.med<-bias.med/true.ab
# efficiency
ab.sd<-stats$SD[2]
```

```
# Combine chains for HPD
if(n.chains>1){
  str(draws.from.bugs)
  draws.to.analyze.as.one.list <-
as.mcmc(do.call(rbind,draws.from.bugs))
  str(draws.to.analyze.as.one.list)
} # closes if n.chains > 1

if(n.chains==1){
  draws.to.analyze.as.one.list <- as.mcmc(draws.from.bugs.as.matrix)
}

# HPDs for each parameter
probability.for.HPD=.95
HPD.interval <- HPDinterval(draws.to.analyze.as.one.list,
prob=probability.for.HPD)
ab.low.hpd <- HPD.interval[2,1]
ab.up.hpd <- HPD.interval[2,2]

# interval outcomes
pow <- NA
if (ab.low.hpd > 0 & ab.up.hpd > 0) {
  pow <- 1
} else {
  pow <- 0
}
coverage <- NA
if (true.ab > ab.low.hpd & true.ab < ab.up.hpd) {
  coverage <- 1
} else {
  coverage<-0
}
typei <- NA
imb <- NA
imb.r <- NA
imb.l <- NA
  if (true.ab > ab.up.hpd) {
    imb.r <- 1
  } else {
    imb.r <- 0
  }
if (true.ab < ab.low.hpd) {
  imb.l <- 1
} else {
  imb.l <- 0
}
imb <- imb.r-imb.l
int.width <- ab.up.hpd - ab.low.hpd
try.mean<- mean(raw.data$x1)

sum.stats <- cbind(which.rep,ab.mean, ab.med, ab.low.hpd,
ab.up.hpd,try.mean,
                   bias.mean, bias.med, ab.sd, rel.bias.mean,
rel.bias.med,
```

```
                    pow, typei, coverage, imb, int.width
)

sum.stats.name <- paste("sum.stats",which.rep,sep="")
assign(sum.stats.name, sum.stats)
names(sum.stats) <- sum.stats.name

sum.stats.all[which.rep, ] <- sum.stats
rm(model.in.winbugs)
rm(sum.stats)
} # closes loop over replications

# Write out the summary statistics

write.csv(
  x=sum.stats.all,
  file="Summary Statistics Diffuse Conjugate.csv"
)
```

Analysis using Bayesian methods with accurate informative priors with $N_{prior} = 100$

```
# Call the 'R2WinBUGS' library

library(R2WinBUGS)

# Loop over replications and generate the data

which.rep=0
which.rep=which.rep+1

sum.stats.all = data.frame(which.rep=rep(0, n.reps),
ab.mean=rep(0,n.reps), ab.med=rep(0,n.reps),
                           ab.low.hpd=rep(0,n.reps),
ab.up.hpd=rep(0,n.reps), try.mean=rep(0,n.reps),
                           bias.mean=rep(0,n.reps),
bias.med=rep(0,n.reps), ab.sd=rep(0,n.reps),
                           rel.bias.mean=rep(0,n.reps),
rel.bias.med=rep(0,n.reps),
                           pow=rep(0,n.reps), typei=rep(0,n.reps),
coverage=rep(0,n.reps),
                           imb=rep(0,n.reps), int.width=rep(0,n.reps))


for(which.rep in 1:n.reps){


# Read in the data

  if(1==1){
    file.name <- "Read in and Separate Data.R"
    source(paste(drive.letter, file.name, sep=""))
  }

# Define the model
# by creating a string that is the BUGS code
```

```
modelstring <- as.character("

model{

# Prior distributions

# Measurement model

lam.x2 ~ dnorm(1, 84.9087);
lam.x3 ~ dnorm(1, 84.9087);

lam.m2 ~ dnorm(1, 84.9087);
lam.m3 ~ dnorm(1, 84.9087);

lam.y2 ~ dnorm(1, 84.9087);
lam.y3 ~ dnorm(1, 84.9087);

tau.x1 ~ dgamma(50, 21.43);
tau.x2 ~ dgamma(50, 21.43);
tau.x3 ~ dgamma(50, 21.43);

sigma.x1 <-1/tau.x1;
sigma.x2 <-1/tau.x2;
sigma.x3 <-1/tau.x3;

tau.m1 ~ dgamma(50, 21.43);
tau.m2 ~ dgamma(50, 21.43);
tau.m3 ~ dgamma(50, 21.43);

sigma.m1 <-1/tau.m1;
sigma.m2 <-1/tau.m2;
sigma.m3 <-1/tau.m3;

tau.y1 ~ dgamma(50, 21.43);
tau.y2 ~ dgamma(50, 21.43);
tau.y3 ~ dgamma(50, 21.43);

sigma.y1 <-1/tau.y1;
sigma.y2 <-1/tau.y2;
sigma.y3 <-1/tau.y3;

# Structural model

tau.latx ~ dgamma(50, 50);
tau.res.latm ~ dgamma(50, 32);
tau.res.laty ~ dgamma(50, 45.84);

a ~ dnorm(0.6, 80.20726);
b ~ dnorm(0.2, 45.27113);
cp ~ dnorm(0.12, 45.72357);

ab <-a*b

# Conditional probability of the data
```

```
        for(i in 1:n){
    x1.mean[i] <- 1*Ksi1[i];
                x2.mean[i] <- lam.x2*Ksi1[i];
                x3.mean[i] <- lam.x3*Ksi1[i];

    x4.mean[i] <- 1*Ksi2[i];
                x5.mean[i] <- lam.m2*Ksi2[i];
                x6.mean[i] <- lam.m3*Ksi2[i];

    x7.mean[i] <- 1*Ksi3[i];
                x8.mean[i] <- lam.y2*Ksi3[i];
                x9.mean[i] <- lam.y3*Ksi3[i];

                Ksi1[i] ~ dnorm(0,tau.latx)

        x1[i] ~ dnorm(x1.mean[i], tau.x1);
                x2[i] ~ dnorm(x2.mean[i], tau.x2);
                x3[i] ~ dnorm(x3.mean[i], tau.x3);

    x4[i] ~ dnorm(x4.mean[i], tau.m1);
        x5[i] ~ dnorm(x5.mean[i], tau.m2);
                x6[i] ~ dnorm(x6.mean[i], tau.m3);

    x7[i] ~ dnorm(x7.mean[i], tau.y1);
    x8[i] ~ dnorm(x8.mean[i], tau.y2);
                x9[i] ~ dnorm(x9.mean[i], tau.y3);

    Ksi2.mean[i] <- a*Ksi1[i];
                Ksi3.mean[i] <- cp*Ksi1[i] + b*Ksi2[i];

  Ksi2[i] ~ dnorm(Ksi2.mean[i],tau.res.latm)
  Ksi3[i] ~ dnorm(Ksi3.mean[i],tau.res.laty)
        }
}

")

# Write out the BUGS code to a file

BUGS.code.file.name <- "Accurate.txt"
write(modelstring, BUGS.code.file.name)
write(modelstring, "temp.bug")

# Define data to give to BUGS

raw.data = read.csv(data.file.name)
n <- nrow(raw.data)
x1 = raw.data$x1
x2 = raw.data$x2
x3 = raw.data$x3
x4 = raw.data$x4
x5 = raw.data$x5
x6 = raw.data$x6
x7 = raw.data$x7
```

```
x8 = raw.data$x8
x9 = raw.data$x9

data <- list("n", "x1", "x2", "x3", "x4", "x5", "x6", "x7", "x8", "x9")

# Define parameters to monitor in BUGS

parameters <- c("lam.x2", "lam.x3", "tau.x1", "tau.x2", "tau.x3",
                "lam.m2", "lam.m3", "tau.m1", "tau.m2", "tau.m3",
                "lam.y2", "lam.y3", "tau.y1", "tau.y2", "tau.y3",
                "tau.latx", "tau.res.latm", "tau.res.laty", "a", "b",
"cp", "ab")

# Define initial values to give to BUGS for each of 3 chains

lam.x2.inits.1 = 0
lam.x3.inits.1 = 0.3
tau.x1.inits.1 = 2
tau.x2.inits.1 = 1
tau.x3.inits.1 = .5

lam.m2.inits.1 = 0
lam.m3.inits.1 = 0.3
tau.m1.inits.1 = 2
tau.m2.inits.1 = 1
tau.m3.inits.1 = .5

lam.y2.inits.1 = 0
lam.y3.inits.1 = 0.3
tau.y1.inits.1 = 2
tau.y2.inits.1 = 1
tau.y3.inits.1 = .5

tau.latx.inits.1= .5
tau.res.latm.inits.1= .5
tau.res.laty.inits.1= .5
a.inits.1= .3
b.inits.1= .4
cp.inits.1= .7

inits1 <- list(lam.x2=lam.x2.inits.1, lam.x3=lam.x3.inits.1,
               tau.x1=tau.x1.inits.1, tau.x2=tau.x2.inits.1,
tau.x3=tau.x3.inits.1,
               lam.m2=lam.m2.inits.1, lam.m3=lam.m3.inits.1,
               tau.m1=tau.m1.inits.1, tau.m2=tau.m2.inits.1,
tau.m3=tau.m3.inits.1,
               lam.y2=lam.y2.inits.1, lam.y3=lam.y3.inits.1,
               tau.y1=tau.y1.inits.1, tau.y2=tau.y2.inits.1,
tau.y3=tau.y3.inits.1,
               tau.latx=tau.latx.inits.1,
tau.res.latm=tau.res.latm.inits.1,
               tau.res.laty=tau.res.laty.inits.1, a=a.inits.1,
b=b.inits.1, cp=cp.inits.1)

lam.x2.inits.2 = 0.3
```

```
lam.x3.inits.2 = 0.1
tau.x1.inits.2 = .5
tau.x2.inits.2 = 2
tau.x3.inits.2 = 1

lam.m2.inits.2 = 0.3
lam.m3.inits.2 = 0
tau.m1.inits.2 = 1
tau.m2.inits.2 = .5
tau.m3.inits.2 = 2

lam.y2.inits.2 = 0.3
lam.y3.inits.2 = 0
tau.y1.inits.2 = 1
tau.y2.inits.2 = .5
tau.y3.inits.2 = 2

tau.latx.inits.2= 1
tau.res.latm.inits.2= 1
tau.res.laty.inits.2= 1
a.inits.2= .7
b.inits.2= .3
cp.inits.2= .4

inits2 <- list(lam.x2=lam.x2.inits.2, lam.x3=lam.x3.inits.2,
               tau.x1=tau.x1.inits.2, tau.x2=tau.x2.inits.2,
tau.x3=tau.x3.inits.2,
               lam.m2=lam.m2.inits.2, lam.m3=lam.m3.inits.2,
               tau.m1=tau.m1.inits.2, tau.m2=tau.m2.inits.2,
tau.m3=tau.m3.inits.2,
               lam.y2=lam.y2.inits.2, lam.y3=lam.y3.inits.2,
               tau.y1=tau.y1.inits.2, tau.y2=tau.y2.inits.2,
tau.y3=tau.y3.inits.2,
               tau.latx=tau.latx.inits.2,
tau.res.latm=tau.res.latm.inits.2,
               tau.res.laty=tau.res.laty.inits.2, a=a.inits.2,
b=b.inits.2, cp=cp.inits.2)

lam.x2.inits.3 = 0.01
lam.x3.inits.3 = 0
tau.x1.inits.3 = 1
tau.x2.inits.3 = .5
tau.x3.inits.3 = 2

lam.m2.inits.3 = 0.3
lam.m3.inits.3 = 0.7
tau.m1.inits.3 = 2.4
tau.m2.inits.3 = 1.7
tau.m3.inits.3 = .2

lam.y2.inits.3 = 0.3
lam.y3.inits.3 = 3
tau.y1.inits.3 = 2.5
tau.y2.inits.3 = 1.7
tau.y3.inits.3 = .85
```

```
tau.latx.inits.3= .15
tau.res.latm.inits.3= .75
tau.res.laty.inits.3= 3.5
a.inits.3= 2.3
b.inits.3= 1.4
cp.inits.3= .2


inits3 <- list(lam.x2=lam.x2.inits.3, lam.x3=lam.x3.inits.3,
               tau.x1=tau.x1.inits.3, tau.x2=tau.x2.inits.3,
tau.x3=tau.x3.inits.3,
               lam.m2=lam.m2.inits.3, lam.m3=lam.m3.inits.3,
               tau.m1=tau.m1.inits.3, tau.m2=tau.m2.inits.3,
tau.m3=tau.m3.inits.3,
               lam.y2=lam.y2.inits.3, lam.y3=lam.y3.inits.3,
               tau.y1=tau.y1.inits.3, tau.y2=tau.y2.inits.3,
tau.y3=tau.y3.inits.3,
               tau.latx=tau.latx.inits.3,
tau.res.latm=tau.res.latm.inits.3,
               tau.res.laty=tau.res.laty.inits.3, a=a.inits.3,
b=b.inits.3, cp=cp.inits.3)


inits <- list(inits1, inits2, inits3)

# Choose
# the number of chains
# the number of iterations to burn-in,
# the number of iterations to thin by,
# the total number of iterations

n.chains = 3
n.burnin = 1500
n.thin = 1
n.iters.total.per.chain = 5000

# Call WinBUGS to run the model, returning a BUGS object

model.in.winbugs <- bugs(
  data=data,
  inits=inits,
  parameters.to.save=parameters,
  model.file=BUGS.code.file.name,
  n.chains=n.chains,
  n.iter=n.iters.total.per.chain,
        n.burnin=n.burnin,
        n.thin=n.thin,
        debug=FALSE,
        codaPkg=TRUE,
  working.directory=getwd(),
  bugs.directory = 'C:\\Program Files\\WinBUGS14',
        DIC=FALSE,
)
```

```
# This code reads in the draws from BUGS

library(coda)

# Define the R object with the names of the coda files

coda.file.names <- model.in.winbugs

# Read the files into coda

draws.from.bugs <- read.bugs(coda.file.names)

# Convert the draws to a matrix

draws.from.bugs.as.matrix <- as.matrix(draws.from.bugs)

# This code summarizes the draws
# Combine chains for summaries

coda.options(combine.stats=TRUE, combine.plots=TRUE)

# Extract the summary statistics
#   Usual
#   Percentiles
#   HPD

# Exporting the iterations

export.draws <- do.call(rbind.data.frame, draws.from.bugs)
export.draws.name <- paste("accurate_both_n100_draws.", which.rep,
".csv", sep="")

write.csv(x=export.draws,
  file=export.draws.name,
  row.names=FALSE,
  col.names=TRUE)

summary.stats <- summary(draws.from.bugs)

stats<- as.data.frame(summary.stats$statistics)
ab.mean <- stats$Mean[2] # this is the post mean of ab

quant <-as.data.frame(summary.stats$quantiles)

ab.med <- quant[2,3] #post median of ab

true.ab<-a*b

# bias
bias.mean<-ab.mean-true.ab
bias.med<-ab.med-true.ab

# relative bias
rel.bias.mean<-bias.mean/true.ab
rel.bias.med<-bias.med/true.ab
```

```
# efficiency
ab.sd<-stats$SD[2]

# Combine chains for HPD

if(n.chains>1){
  str(draws.from.bugs)
  draws.to.analyze.as.one.list <-
as.mcmc(do.call(rbind,draws.from.bugs))
  str(draws.to.analyze.as.one.list)
} # closes if n.chains > 1

if(n.chains==1){
  draws.to.analyze.as.one.list <- as.mcmc(draws.from.bugs.as.matrix)
}

# HPDs for each parameter

probability.for.HPD=.95
HPD.interval <- HPDinterval(draws.to.analyze.as.one.list,
prob=probability.for.HPD)
ab.low.hpd <- HPD.interval[2,1]
ab.up.hpd <- HPD.interval[2,2]

# interval outcomes

pow <- NA
if (ab.low.hpd > 0 & ab.up.hpd > 0) {
  pow <- 1
} else {
  pow <- 0
}
coverage <- NA
if (true.ab > ab.low.hpd & true.ab < ab.up.hpd) {
  coverage <- 1
} else {
  coverage<-0
}
typei <- NA
imb <- NA
imb.r <- NA
imb.l <- NA
  if (true.ab > ab.up.hpd) {
    imb.r <- 1
  } else {
    imb.r <- 0
  }
if (true.ab < ab.low.hpd) {
  imb.l <- 1
} else {
  imb.l <- 0
}
imb <- imb.r-imb.l
int.width <- ab.up.hpd - ab.low.hpd
```

```
try.mean<- mean(raw.data$x1)

sum.stats <- cbind(which.rep,ab.mean, ab.med, ab.low.hpd,
ab.up.hpd,try.mean,
                   bias.mean, bias.med, ab.sd, rel.bias.mean,
rel.bias.med,
                   pow, typei, coverage, imb, int.width
)

sum.stats.name <- paste("sum.stats",which.rep,sep="")
assign(sum.stats.name, sum.stats)
names(sum.stats) <- sum.stats.name

sum.stats.all[which.rep, ] <- sum.stats
rm(model.in.winbugs)
rm(sum.stats)
} # closes loop over replications

# Write out the summary statistics

write.csv(
  x=sum.stats.all,
  file="Summary Statistics Comb1 N100 Accurate Both.csv"
)
```