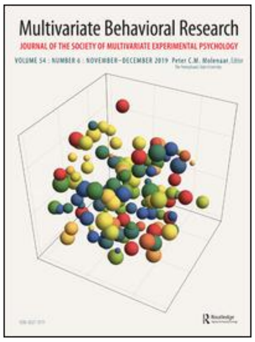

Routledge
Taylor & Francis Group

Multivariate Behavioral Research

ISSN: 0027-3171 (Print) 1532-7906 (Online) Journal homepage: https://www.tandfonline.com/loi/hmbr20

Different Roles of Prior Distributions in the Single Mediator Model with Latent Variables

Milica Miočević, Roy Levy & David P. MacKinnon

To cite this article: Milica Miočević, Roy Levy & David P. MacKinnon (2020): Different Roles of Prior Distributions in the Single Mediator Model with Latent Variables, Multivariate Behavioral Research, DOI: 10.1080/00273171.2019.1709405

To link to this article: https://doi.org/10.1080/00273171.2019.1709405

Published online: 31 Jan 2020.

MULTIVARIATE BEHAVIORAL RESEARCH
https://doi.org/10.1080/00273171.2019.1709405

Different Roles of Prior Distributions in the Single Mediator Model with Latent Variables

Milica Miočević[a], Roy Levy[b], and David P. MacKinnon[c]

[a]Department of Psychology, McGill University, Montreal, Canada; [b]T. Denny Sanford School of Social & Family Dynamics, Arizona State University, Tempe, AZ, USA; [c]Department of Psychology, Arizona State University, Tempe, AZ, USA

ABSTRACT

In manifest variable models, Bayesian methods for mediation analysis can have better statistical properties than commonly used frequentist methods. However, with latent variables, Bayesian mediation analysis with diffuse priors can yield worse statistical properties than frequentist methods, and no study to date has evaluated the impact of informative priors on statistical properties of point and interval summaries of the mediated effect. This article describes the first examination of using fully conjugate and informative (accurate and inaccurate) priors in Bayesian mediation analysis with latent variables. Results suggest that fully conjugate priors and informative priors with the same relative prior sample sizes have notably different effects at $N=200$ and 400, than at $N=50$ and 100. Consequences of a small amount of inaccuracy in priors for loadings can be alleviated by making the prior less informative, whereas the same is not always true of inaccuracy in priors for structural paths. Finally, the consequences of using informative priors depend on the inferential goals of the analysis: inaccurate priors are more detrimental for accurately estimating the mediated effect than for evaluating whether the mediated effect is nonzero. Recommendations are provided about when to gainfully employ Bayesian mediation analysis with latent variables.

KEYWORDS

Bayesian; mediation analysis; latent variable model; informative priors

Introduction

Mediation analysis is used to study intermediate variables that transmit the effect of an independent variable to a dependent variable. Notable developments in mediation analysis in the past decades consisted of identifying optimal ways for testing the significance of the mediated effect (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; MacKinnon, Lockwood, & Williams, 2004; Shrout & Bolger, 2002), the extension of mediation analysis to more complex models (Finch, West, & MacKinnon, 1997; Krull & MacKinnon, 1999; Preacher, Rucker, & Hayes, 2007), the definition of conditions for causal inferences in mediation analysis (Coffman & Zhong, 2012; Imai, Keele, & Tingley, 2010; Jo, Stuart, MacKinnon, & Vinokur, 2011; Maxwell & Cole, 2007; Valeri & VanderWeele, 2013), and the description and implementation of mediation analysis in the Bayesian framework (Enders, Fairchild, & MacKinnon, 2013; Yuan & MacKinnon, 2009).

Bayesian methods with informative priors may be used to increase power to detect the mediated effect in models with manifest variables (Miočević, MacKinnon, & Levy, 2017; Yuan & MacKinnon, 2009). In the absence of relevant prior information, posterior summaries of the mediated effect obtained using Bayesian methods with diffuse priors have comparable or better (Koopman, Howe, Hollenbeck, & Sin, 2015; Miočević et al., 2017) statistical properties than commonly used frequentist methods, but have the advantage of probabilistic interpretations (Little, 2006). More specifically, diffuse (noninformative) prior distributions are used to communicate ignorance about the sign and magnitude of coefficients in the mediation model; the most common choice for diffuse priors for coefficients in mediation analysis are normal distributions centered at zero with a large variance (Koopman et al., 2015; Miočević et al., 2017; Yuan & MacKinnon, 2009). Bayesian methods with diffuse prior distributions have higher power than normal theory confidence limits because they do not require any assumptions about the distribution of the mediated effect, and normal theory confidence limits assume that the distribution of the

CONTACT Milica Miočević milica.miocevic@mcgill.ca Department of Psychology, McGill University, Montreal, Canada.

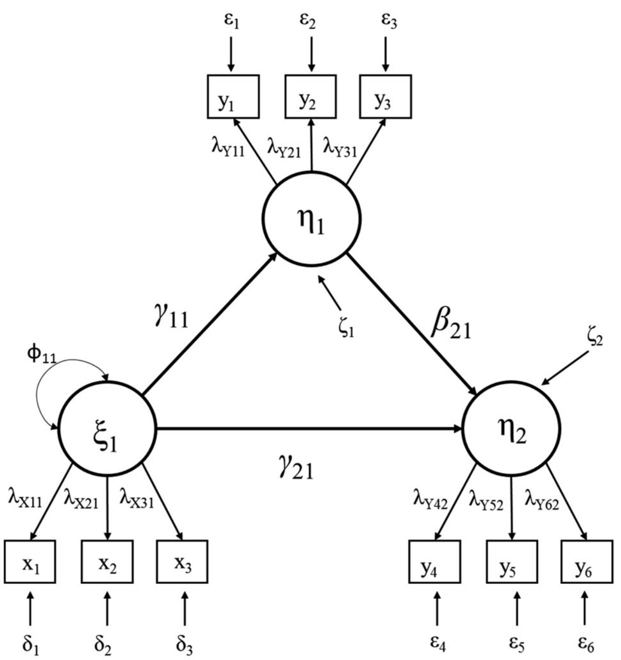

Figure 1. Single mediator model with latent variables and three manifest indicators per latent variable.

mediated effect is symmetric, which is often not the case for the distribution of the product (Craig, 1936; Lomnicki, 1967; Springer & Thompson, 1966). Furthermore, Bayesian methods with diffuse priors do not encounter as many instances of Type I error rates as the bias-corrected bootstrap (Koopman et al., 2015), which is another commonly used frequentist method, despite the clear warnings in the literature about when bias-corrected bootstrap confidence limits for the mediated effect have excessive Type I error rates (Fritz, Taylor, & MacKinnon, 2012).

Bayesian mediation analysis has been extensively studied for manifest variable models (Chen, Choi, Weiss, & Stapleton, 2014; Enders et al., 2013; Miočević et al., 2017; Yuan & MacKinnon, 2009), and the examination of Bayesian methods with diffuse priors for latent variable mediation models has just begun (Chen, Choi, Weiss, & Stapleton, 2014; van Erp, Mulder, & Oberski, 2018). This article contains simulation studies evaluating the following two questions related to Bayesian mediation analysis with latent variables: (1) What are the statistical properties of Bayesian point and interval summaries of the mediated effect when the researcher selects priors that communicate ignorance about the sign and magnitude of the structural paths and loadings?; 2) Can the negative impact of inaccurate priors for loadings and structural paths be alleviated by making the prior less informative, and are the effects the same on statistical properties of point and interval summaries of the mediated effect? The results of the simulation studies in this article present a more nuanced picture of the effects that different assumptions and levels of informativeness encoded in prior distributions have on statistical properties of the mediated effect in the single mediator model with latent variables. The following sections describe the single mediator model with latent variables, mention considerations when fitting this model in the Bayesian framework, offer a description of conjugate priors for model parameters in Bayesian structural equation modeling (SEM), and conclude with guiding questions for the simulation studies.

Single mediator model with latent variables

The single mediator model with latent variables and three indicators per latent variable, as described by Finch et al. (1997), consists of a measurement model for each latent variable, and a structural model for the independent variable, mediator, and outcome (Figure 1).

The model is described using Equations (1–3):

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \lambda_{X11} \\ \lambda_{X21} \\ \lambda_{X31} \end{bmatrix} \xi_1 + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix} \tag{1}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} = \begin{bmatrix} \lambda_{Y11} & 0 \\ \lambda_{Y21} & 0 \\ \lambda_{Y31} & 0 \\ 0 & \lambda_{Y42} \\ 0 & \lambda_{Y52} \\ 0 & \lambda_{Y62} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix} \tag{2}$$

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \beta_{21} & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} \\ \gamma_{21} \end{bmatrix} \xi_1 + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}. \tag{3}$$

The mediated effect is computed as the product of structural paths $\gamma_{11}\beta_{21}$. There are at least two possible inferential goals[1] in mediation analysis: to estimate the value of the indirect (mediated) effect and to evaluate whether the indirect effect is different from zero. The first inferential goal deals with the point estimate/summary of the mediated effect and the second inferential goal may be accomplished by evaluating whether the interval estimate/summary of the mediated effect includes zero. Previous studies of the statistical properties of the point and interval estimates of the mediated effect in latent variable models found that ML point estimates of the indirect and direct effects have less than 10% relative bias for sample sizes of 150, 250, 500, and 1000 (Finch et al., 1997).

[1]There are other inferential goals of mediation analysis, e.g., researchers could be interested if the direct effect is zero or not, thus evaluating whether the effect is "fully" or "partially" mediated. In this manuscript, the two inferential goals we mention in the main text are the only goals of interest for this study.

Out of 11 estimators of the mediated effect in latent variable models, the likelihood-based confidence intervals, the percentile bootstrap, and the distribution of the product method emerged as best methods based on power, Type I error rates, and 95% coverage for sample sizes of 100, 200, and 500 (Falk & Biesanz, 2015). Recently, methodologists started examining the statistical properties of Bayesian methods for the single mediator model with latent variables estimated as a Bayesian SEM with continuous (Chen et al., 2014) and ordinal (Chen, Zhang, & Choi, 2015) indicators. Chen et al. (2014) examined the bias of the posterior mean and coverage of equal-tail credibility intervals for the mediated effect obtained using diffuse uniform priors for loadings and structural paths in the single mediator model with latent variables, continuous indicators, and complete mediation at $N=50$, 100, and 400. Findings indicate that ML can encounter nonconvergence issues at $N=50$ and 100. However, in combinations where ML converged, the posterior mean had more instances of relative bias above 10% than the ML estimate. In general, the posterior mean had higher relative bias and was more efficient (i.e., had lower root mean-squared error (RMSE) values) than the ML estimate, and equal-tail credibility intervals had more instances of miscoverage (coverage both above and below the nominal level) than bias-corrected bootstrap confidence limits. Thus, even though Bayesian methods offer a solution when ML does not converge, the posterior mean obtained with uninformative nonconjugate priors for structural paths and loadings is generally more biased than the ML point estimate at sample sizes up to 100.

Fitting the single mediator model with latent variables as a Bayesian SEM

For the sake of brevity, this section covers only the considerations specific to fitting a single mediator model with latent variables in the Bayesian SEM framework. For details on how to fit this model demonstrated through an empirical example, see Miočević (2019). For a general introduction to Bayesian methods for social sciences, see papers by van de Schoot et al. (2014) and van de Schoot and Depaoli (2014). For introductions to Bayesian SEM see chapters by Kaplan and Depaoli (2012) and Levy and Choi (2013). For information on Bayesian mediation analysis with manifest variables see papers by Yuan and MacKinnon (2009), Enders et al. (2013), and Miočević et al. (2017).

Fitting the single mediator model with latent variables as a Bayesian SEM requires specifying prior distributions for all unknown parameters. In manifest variable mediation models, priors can be specified for paths between the independent variable, mediator, and outcome and the residual variances (Yuan & MacKinnon, 2009), or for the covariance matrix of the independent variable, mediator, and outcome (Enders et al., 2013). The method of covariances has the advantage that the degrees of freedom hyperparameter of the inverse-Wishart prior can be interpreted as the sample size of the prior sample. 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 formulate prior expectations about the path between the independent variable and mediator than about the covariance and variance terms needed to compute this path. For this reason, we use the method of coefficients for Bayesian SEM, and not the method of covariances.

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 a constant (often 1), fixing the average of the loadings of the same latent variable to a constant (often 1), or fixing the variances of latent variables to 1 (Little, Slegers, & Card, 2006). In the Bayesian framework, setting one loading to 1 is the most 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. Setting the average of all loadings of the same latent variable to a constant also requires careful consideration when selecting hyperparameters of univariate priors for the loadings. For this reason, the simulation studies in this article use the marker-variable method.

The first step of a Bayesian analysis is the specification of a prior distribution. Note that 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. One possible noninformative 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); this is the definition of uninformativeness for the priors for the variance parameters in this study. In general, priors are labeled as diffuse to indicate that they carry no information, and that we expect such priors to yield Bayesian point summaries that are numerically identical to results from a frequentist analysis. However, whether a prior can be considered diffuse or uninformative depends on the amount of information carried by the data; an inverse-gamma prior with both hyperparameters close to 0 (e.g., equal to 0.001) for a variance parameter can be considered uninformative in some settings, but as Gelman (2006) showed, this prior is not uninformative for a level-2 variance parameter in a multilevel model with only 8 clusters. For more on different kinds of diffuse priors in Bayesian SEM and the statistical properties of the indirect effect obtained using such priors, see van Erp et al. (2018).

Conjugate prior distributions lead to posterior distributions of the same parametric form (Gelman et al., 2013). We distinguish between two kinds of conjugate priors: conditionally conjugate, and fully conjugate. In the *conditionally* conjugate prior specification, normal distributions are specified for loadings and structural paths in SEM (Kaplan & Depaoli, 2012). The conditionally conjugate priors in this study were normal priors for loadings and structural paths centered at 0 with variance hyperparameters that do not depend on the corresponding residual variance (i.e., 1000 and 100); these priors are referred to as "diffuse generic priors" in the remainder of the study.

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 hyperparameters $\alpha = \nu/2$ and $\beta = {\sigma_0}^2\nu/2$, where ν denotes the degrees of freedom parameter of the inverse-χ^2 and ${\sigma_0}^2$ denotes the scale of σ^2) and the conditional distribution of μ given σ^2 is normal (Gelman et al., 2013).

Recall that the conjugate prior for the variance is an inverse-gamma distribution (Gelman et al., 2013). Thus, the fully conjugate priors for loading and structural paths are normal distributions conditional on the corresponding residual variances, which are assigned inverse-gamma priors. The fully conjugate priors specified in this project were normal priors for loadings and structural paths centered at zero with a variance hyperparameter equal to the corresponding residual variance, and inverse-gamma priors with hyperparameters $\alpha = \beta = 0.5$ for measurement error variances of manifest indicators, residual variances of endogenous latent variables, and the variances of the exogenous latent variable. Unlike normal distributions where the informativeness of the prior is determined by the variance hyperparameter, inverse-gamma distributions allow for encoding the desired weight of the prior information in terms of prior sample size. Given the relationship between the scaled inverse-chi-square distribution and the inverse-gamma distribution, the shape (α) hyperparameter of the inverse-gamma prior can be thought of as half of the prior sample size, and scale (β) hyperparameter can be thought of as the product of half of the prior sample size and the observed value of the corresponding variance in the prior sample (Gelman et al., 2013). Thus, the inverse-gamma priors with both hyperparameters equal to 0.5 encode the assumption that the best guess for the prior variance is 1, and that the best guess carries the weight of one observation. When the variance parameter being conditioned on is large, the fully conjugate specification will lead to a diffuse prior. However, this was not the case in the simulation studies in this article because both the measurement error variances and the residual variances of latent variables are generated to be smaller than 1, thus leading to informative priors when the variance hyperparameter of normal priors is set equal to either the measurement error variance or the residual variance of either η_1 or η_2 (i.e., the latent mediator and latent outcome variables from Equation (2), respectively). Note that both fully conjugate and diffuse generic priors in this study encode that the researcher's best guess about each freely estimated loading and structural path is zero. Thus, both of these priors allow the researcher to avoid making any guesses about the signs of these parameters; however, in the simulation studies for this article, fully conjugate priors are notably more informative than diffuse generic priors.

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 is done by using Markov chain Monte Carlo (MCMC) methods to approximate the posterior distribution. For an introduction to MCMC estimation, see the paper by Brooks (1998), and for accessible guidelines on how to diagnose convergence of MCMC see the paper by Sinharay (2004). The mediated effect is computed as the product of paths $\gamma_{11}\beta_{21}$ at each MCMC iteration, thus yielding an approximation of the marginal posterior of the mediated effect. The posterior for the mediated effect can be summarized using the mean, median, and mode as

Table 1. Parameter combinations for the Monte Carlo studies.

Combination	γ_{11}	β_{21}	γ_{21}
1	0.60	0.20	0.12
2	0.30	0.40	0.12
3	0.30	0.40	0.36
4	0	0.40	0.12
5	0.40	0	0.12
6	0	0	0.12

point summaries, and equal-tail and highest posterior density (HPD) intervals as interval summaries. HPD intervals have the property that no value outside of the interval has higher probability than values inside of the interval (Gelman et al., 2013). The distribution of the product of two normal variates is often asymmetric (Craig, 1936; Lomnicki, 1967; Springer & Thompson, 1966) and methods that take this distribution into account or make no distributional assumptions yield intervals for the mediated effect with better statistical properties than normal theory confidence intervals (Cheung, 2007, 2009; MacKinnon, Fritz, Williams, & Lockwood, 2007; MacKinnon et al., 2002; 2004; Shrout & Bolger, 2002; Tofighi & MacKinnon, 2011; Valente, Gonzalez, Miočević, & MacKinnon, 2016). In the frequentist framework, the distribution of the product confidence limits, obtained analytically using critical values from the distribution of the product of two normal variates, are among the best interval estimators for the mediated effect. This simulation study uses the distribution of the product confidence limits as the frequentist interval estimator and uses only HPD intervals as a Bayesian interval estimator of the mediated effect because HPD intervals are better suited for representing the most probable values in nonsymmetrical posterior distributions.

Goals of the Monte Carlo studies

It is still unclear whether researchers should opt for Bayesian methods with conjugate priors in the absence of prior information about the model parameters. Furthermore, no study to date has evaluated the statistical properties of the point summaries of the indirect effect obtained using fully conjugate priors, and there has not been any research done about the impact of informative priors on the results of Bayesian analyses of the latent variable mediation model.

The Monte Carlo studies in this article were designed to answer the following sets of questions:

Study 1. When there is no prior information on the sign of structural paths and loadings, should researchers use Bayesian methods with fully conjugate priors, Bayesian methods with diffuse generic priors, or ML estimation and the distribution of the product to compute point and interval estimates/summaries of the mediated effect? In other words, how well do Bayesian methods with diffuse generic priors and fully conjugate priors perform in terms of evaluating the magnitude of the indirect effect and assessing if the indirect effect is zero relative to ML point estimates and distribution of the product confidence intervals with sample sizes of 50, 100, 200, and 400?

Study 2. Accurate informative priors will improve statistical properties of point and interval summaries of the mediated effect, whereas inaccurate priors will make the statistical properties worse. However, the consequences of inaccuracy in one's prior beliefs about a parameter can sometimes be alleviated by making the prior less informative (Depaoli, 2014). Study 2 evaluated 1) the amount of improvement in statistical properties that can be achieved with accurate priors that carry $1/2$ and $1/4$ of the weight of the likelihood function at $N = 50$, 100, 200, and 400 relative to results obtained using ML estimation and the distribution of the product, and 2) whether the negative consequences of inaccuracy in the priors for loadings and structural paths can be alleviated by making the prior carry $1/4$ instead of $1/2$ of the weight of the likelihood function.

Methods

The Monte Carlo studies were carried out using WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000), R (R Core Team, 2014), and R packages lavaan (Rosseel, 2012), RMediation (Tofighi & MacKinnon, 2011), R2WinBUGS (Sturtz, Ligges, & Gelman, 2005), and coda (Plummer, Best, Cowles, & Vines, 2006).[2] A program was written to draw and store 500 samples from the six populations based on parameter values in Combinations 1–6 (Table 1). The parameter values for structural parameters in Combinations 1–3 were modeled after parameter values in the paper by Finch et al. (1997). Combinations 4–6 were created to evaluate the Type I error rates of methods under examination.

Data were simulated following Equations (1) and (2) for the measurement model, and Equation (3) for the structural model. The disturbances of endogenous

[2]Note that the manuscript discusses normal distributions for parameters and variables using the variance parametrization, however, the software used in the simulation study uses the precision (i.e., inverse of the variance) parametrization. Thus, in the syntax for the simulation, the second parameter in normal priors is the precision, and the (residual) precision parameters in the model are assigned gamma instead of inverse-gamma priors.

Table 2. Prior specifications for Bayesian methods in Study 1.

	Diffuse generic priors	Fully conjugate priors
$N = 50$	$\lambda_i \sim N\ (0,\ 100)$	$\lambda_i \sim N\ (0,\ \sigma_\varepsilon^2)$
	$\gamma_{11},\ \beta_{21},\ \gamma_{21} \sim N\ (0,\ 100)$	$\gamma_{11} \sim N\ (0,\ \psi_{11}),\ \beta_{21},\gamma_{21} \sim N\ (0,\ \psi_{22})$
	$\Phi_{11},\ \psi_{11},\ \psi_{22},\ \sigma_\delta^2,\ \sigma_\varepsilon^2 \sim IG\ (0.5,\ 0.5)$	$\Phi_{11},\ \psi_{11},\ \psi_{22},\ \sigma_\delta^2,\ \sigma_\varepsilon^2 \sim IG\ (.5,\ .5)$
$N = 100$	$\lambda_i \sim N\ (0,\ 100)$	$\lambda_i \sim N\ (0,\ \sigma_\varepsilon^2)$
	$\gamma_{11},\ \beta_{21},\ \gamma_{21} \sim N\ (0,\ 100)$	$\gamma_{11} \sim N\ (0,\ \psi_{11}),\ \beta_{21},\gamma_{21} \sim N\ (0,\ \psi_{22})$
	$\Phi_{11},\ \psi_{11},\ \psi_{22},\ \sigma_\delta^2,\ \sigma_\varepsilon^2 \sim IG\ (0.5,\ 0.5)$	$\Phi_{11},\ \psi_{11},\ \psi_{22},\ \sigma_\delta^2,\ \sigma_\varepsilon^2 \sim IG\ (0.5,\ 0.5)$
$N = 200$	$\lambda_i \sim N\ (0,\ 1000)$	$\lambda_i \sim N\ (0,\ \sigma_\varepsilon^2)$
	$\gamma_{11},\ \beta_{21},\ \gamma_{21} \sim N\ (0,\ 1000)$	$\gamma_{11} \sim N\ (0,\ \psi_{11}),\ \beta_{21},\gamma_{21} \sim N\ (0,\ \psi_{22})$
	$\Phi_{11},\ \psi_{11},\ \psi_{22},\ \sigma_\delta^2,\ \sigma_\varepsilon^2 \sim IG\ (0.001,\ 0.001)$	$\Phi_{11},\ \psi_{11},\ \psi_{22},\ \sigma_\delta^2,\ \sigma_\varepsilon^2 \sim IG\ (0.5,\ 0.5)$
$N = 400$	$\lambda_i \sim N\ (0,\ 1000)$	$\lambda_i \sim N\ (0,\ \sigma_\varepsilon^2)$
	$\gamma_{11},\ \beta_{21},\ \gamma_{21} \sim N\ (0,\ 1000)$	$\gamma_{11} \sim N\ (0,\ \psi_{11}),\ \beta_{21},\gamma_{21} \sim N\ (0,\ \psi_{22})$
	$\Phi_{11},\ \psi_{11},\ \psi_{22},\ \sigma_\delta^2,\ \sigma_\varepsilon^2 \sim IG\ (0.001,\ 0.001)$	$\Phi_{11},\ \psi_{11},\ \psi_{22},\ \sigma_\delta^2,\ \sigma_\varepsilon^2 \sim IG\ (0.5,\ 0.5)$

Note. This table contains specifications of diffuse generic and fully conjugate prior distributions for measurement and structural model parameters in Study 1. The first hyperparameter in the normal priors is the mean, and the second hyperparameter is the variance.

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 variances equal to 1. In all 6 parameter combinations, all loadings were simulated to equal 1, and reliability of 0.7 for each manifest indicator was obtained by simulating the measurement errors of manifest indicators to follow normal distributions with a mean of 0 and variances equal to 0.4286.

The sample sizes of the observed (focal) sample examined in this study were $N = 50$, 100, 200, and 400. In frequentist analyses ML estimation implemented in the R package lavaan (Rosseel, 2012) was used for point estimation, and the distribution of the product confidence limits were obtained using ML estimates of paths γ_{11} and β_{21}, and their standard errors that were entered into the *medci* function in the R package RMediation (Tofighi & MacKinnon, 2011). In the Bayesian analyses the mean, median, and mode of the posterior for the mediated effect were used as point summaries, and HPD intervals were used as interval summaries. The software, sample sizes, parameter values in Combinations 1–6, and Bayesian and frequentist point and interval summaries/estimates were the same in all simulation studies. Each Bayesian analysis featured 1500 burn-in iterations, 3 chains, and 5000 total iterations per chain (3500 iterations post burn-in, thus leading to posteriors based on 10,500 iterations). Criteria used to diagnose convergence were values of the potential scale reduction factor (PSRF; Brooks & Gelman, 1998), where values <1.1 were considered evidence of convergence, and trace plots (good mixing of the three chains was considered evidence of convergence). The thinning period was set to 1 (i.e., no thinning), following the recommendations by Link and Eaton (2012). Subsequent sections only contain information relevant to each specific simulation study.

Study 1

The first study compares frequentist methods to Bayesian methods with diffuse generic priors and fully conjugate priors that had the same mean hyperparameters (i.e., zero) as the diffuse generic priors but different variance hyperparameters (i.e., for loadings the variance hyperparameter was equal to the measurement error variance of the indicator, and for structural paths the variance hyperparameter was equal to the residual variance of the corresponding endogenous latent variable).

A preliminary analysis of a single simulated data set with diffuse generic priors showed that for sample sizes of 200 and 400, the chains converged by 1500 iterations. For sample sizes of 50 and 100, the chains in the analysis with diffuse generic priors did not converge by 1500 iterations like they did for larger sample sizes, and empirical testing showed that the chains did not even converge after 500,000 iterations. The lack of convergence after this many draws suggested that the priors were too diffuse, and thus for $N = 50$ and 100 the diffuse generic priors were made more informative. The best guess for each parameter was still the same as in the diffuse generic priors for $N = 200$ and 400, however, the variance hyperparameter of the normal priors for loadings and structural paths was decreased from 1000 to 100 (thus making the prior 10 times more informative), and for variance and residual variance parameters the best guess was assigned the weight of 1 observation for $N = 50$ and 100 as opposed to 0.002 "observations" for $N = 200$ and 400. The priors for all classes of parameters in Study 1 are shown in Table 2.

All Bayesian analyses in studies 1–3 featured 3 chains, 1500 burn-in iterations, and 3500 iterations post burn-in, thus leading to posteriors based on 10,500 iterations.

Study 2

The second study evaluates the impact of informative (accurate and inaccurate) priors for loadings and structural paths, examines whether the negative consequences of inaccurate priors can be alleviated by making the prior 50% less informative, and reports on the extent to which inaccuracy is detrimental for point summaries *versus* interval summaries of the indirect effect.

Accurate and inaccurate priors were designed to carry 25% and 50% of the information carried by the likelihood function of this study. The choice to have this percentage be below 100% was made to emulate a scenario where the researcher either only has access to published findings from a prior study that had a lower sample size than this study, or the researcher herself chooses not to weigh the published findings from the previous study more heavily than the likelihood function from this study. In this study, all priors for variance (and residual) variance parameters were accurate and had the same level of informativeness as the priors for all other parameters in a given condition. The hyperparameters for the accurate priors for variance parameters were computed analytically by plugging in the true value of the variance term and the size of the prior sample into the formulas for the parameters of the inverse-gamma distribution. For example, the true measurement error variance for indicator x_1 is equal to 0.4286, and the hyperparameters of the inverse-gamma prior for this measurement error variance at $N_{\text{prior}} = 100$ were computed as follows:

$$\theta_{\delta 1} \sim IG\left(\frac{100}{2}, \frac{100}{2}0.4286\right) = IG(50, 21.43).$$

Unlike the inverse-gamma distribution, the normal distribution does not have a spread parameter with a direct interpretation in terms of sample size. Normal priors for structural paths and loadings 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 and 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 and the population covariance matrix as input. 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 variance hyperparameter for the prior being constructed was computed as the squared standard error from the ML results. Mean hyperparameters of accurate priors for structural paths were set equal to the population (i.e., simulated) value of the corresponding structural path. The study examined four different amounts of inaccuracy in the priors for structural paths: 0.5*sd*, 1*sd*, 2*sd*, and 3*sd*. The desired amount of inaccuracy in the mean hyperparameters of normal priors for structural paths was obtained either by adding (Combinations 4–6) or by subtracting (Combinations 1–3) the appropriate number of standard deviations to/from the true value of the corresponding structural path. For example, the inaccurate prior with 2*sd* of inaccuracy for β_{21} at $N_{\text{prior}} = 50$ in Combination 1 was obtained by subtracting $2\sqrt{0.044}$ (two times the square root of the empirically obtained variance at $N_{\text{prior}} = 50$) from the true value 0.20, thus yielding $0.20 - 2\sqrt{0.044} = 0.20 - 2(0.21) = 0.20 - 0.42 = -0.22$ as the mean hyperparameter of the prior for β_{21}. In conditions with inaccurate priors for structural paths, the priors for loadings and variance parameters (i.e., measurement error variances, the variances of the exogenous latent variable, and the residual variances of the endogenous latent variables) remained accurate.

Since composite reliability and its standard error are not obtained in ML estimation of the model, the distribution of composite reliability at N_{prior} = 25, 50, 100, and 200 had to be approximated to compute the standard deviation. This was done by simulating 1000 values of the two freely estimated loadings, the variance of the latent variable, and measurement error variances from their true distributions based on N_{prior} = 25, 50, 100, and 200, and plugging the values into 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

Table 3. Prior specifications for Bayesian methods with *.5sd* inaccuracy in the prior expectations for loadings and structural parameters.

	Inaccurate loadings	Inaccurate structural
	Combination 1 ($\gamma_{11} = 0.60$, $\beta_{21} = 0.20$, $\gamma_{21} = 0.12$)	
$N_{prior} = 25$	$\lambda_i \sim N$ (0.84, 0.047)	$\lambda_i \sim N$ (1, 0.047)
	$\gamma_{11} \sim N$ (0.60, 0.050)	$\gamma_{11} \sim N$ (0.49, 0.050)
	$\beta_{21} \sim N$ (0.20, 0.088)	$\beta_{21} \sim N$ (0.05, 0.088)
	$\gamma_{21} \sim N$ (0.12, 0.087)	$\gamma_{21} \sim N$ (−0.03, 0.087)
$N_{prior} = 50$	$\lambda_i \sim N$ (0.88, 0.024)	$\lambda_i \sim N$ (1, 0.024)
	$\gamma_{11} \sim N$ (0.60, 0.025)	$\gamma_{11} \sim N$ (0.52, 0.025)
	$\beta_{21} \sim N$ (0.20, 0.044)	$\beta_{21} \sim N$ (0.09, 0.044)
	$\gamma_{21} \sim N$ (0.12, 0.044)	$\gamma_{21} \sim N$ (0.01, 0.044)
$N_{prior} = 100$	$\lambda_i \sim N$ (0.91, 0.012)	$\lambda_i \sim N$ (1, 0.012)
	$\gamma_{11} \sim N$ (0.6, 0.012)	$\gamma_{11} \sim N$ (0.54, 0.012)
	$\beta_{21} \sim N$ (0.2, 0.022)	$\beta_{21} \sim N$ (0.13, 0.022)
	$\gamma_{21} \sim N$ (0.12, 0.022)	$\gamma_{21} \sim N$ (0.05, 0.022)
$N_{prior} = 200$	$\lambda_i \sim N$ (0.93, 0.006)	$\lambda_i \sim N$ (1, 0.006)
	$\gamma_{11} \sim N$ (0.6, 0.006)	$\gamma_{11} \sim N$ (0.56, 0.006)
	$\beta_{21} \sim N$ (0.2, 0.011)	$\beta_{21} \sim N$ (0.15, 0.011)
	$\gamma_{21} \sim N$ (.12, .011)	$\gamma_{21} \sim N$ (0.07, 0.011)
	Combination 2 ($\gamma_{11} = 0.30$, $\beta_{21} = 0.40$, $\gamma_{21} = 0.12$)	
$N_{prior} = 25$	$\lambda_i \sim N$ (0.84, 0.047)	$\lambda_i \sim N$ (1, 0.047)
	$\gamma_{11} \sim N$ (0.30, 0.052)	$\gamma_{11} \sim N$ (0.19, 0.052)
	$\beta_{21} \sim N$ (0.40, 0.055)	$\beta_{21} \sim N$ (0.28, 0.055)
	$\gamma_{21} \sim N$ (0.12, 0.050)	$\gamma_{21} \sim N$ (0.01, 0.050)
$N_{prior} = 50$	$\lambda_i \sim N$ (0.88, 0.024)	$\lambda_i \sim N$ (1, 0.024)
	$\gamma_{11} \sim N$ (0.30, 0.026)	$\gamma_{11} \sim N$ (0.22, 0.026)
	$\beta_{21} \sim N$ (0.40, 0.027)	$\beta_{21} \sim N$ (0.32, 0.027)
	$\gamma_{21} \sim N$ (0.12, 0.025)	$\gamma_{21} \sim N$ (0.04, 0.025)
$N_{prior} = 100$	$\lambda_i \sim N$ (0.91, 0.012)	$\lambda_i \sim N$ (1, 0.012)
	$\gamma_{11} \sim N$ (0.3, 0.013)	$\gamma_{11} \sim N$ (0.24, 0.013)
	$\beta_{21} \sim N$ (0.4, 0.014)	$\beta_{21} \sim N$ (0.34, 0.014)
	$\gamma_{21} \sim N$ (0.12, 0.012)	$\gamma_{21} \sim N$ (0.06, 0.012)
$N_{prior} = 200$	$\lambda_i \sim N$ (0.93, 0.006)	$\lambda_i \sim N$ (1, 0.006)
	$\gamma_{11} \sim N$ (0.3, 0.006)	$\gamma_{11} \sim N$ (0.26, 0.006)
	$\beta_{21} \sim N$ (0.4, 0.007)	$\beta_{21} \sim N$ (0.36, 0.007)
	$\gamma_{21} \sim N$ (0.12, 0.006)	$\gamma_{21} \sim N$ (0.08, 0.006)
	Combination 3 ($\gamma_{11} = 0.30$, $\beta_{21} = 0.40$, $\gamma_{21} = 0.36$)	
$N_{prior} = 25$	$\lambda_i \sim N$ (0.84, 0.047)	$\lambda_i \sim N$ (1, 0.047)
	$\gamma_{11} \sim N$ (0.30, 0.052)	$\gamma_{11} \sim N$ (0.18, 0.052)
	$\beta_{21} \sim N$ (0.40, 0.047)	$\beta_{21} \sim N$ (0.29, 0.047)
	$\gamma_{21} \sim N$ (0.36, 0.046)	$\gamma_{21} \sim N$ (0.25, 0.046)
$N_{prior} = 50$	$\lambda_i \sim N$ (0.88, 0.024)	$\lambda_i \sim N$ (1, 0.024)
	$\gamma_{11} \sim N$ (0.30, 0.026)	$\gamma_{11} \sim N$ (0.22, 0.026)
	$\beta_{21} \sim N$ (.40, .023)	$\beta_{21} \sim N$ (0.32, 0.023)
	$\gamma_{21} \sim N$ (0.36, 0.023)	$\gamma_{21} \sim N$ (0.28, 0.023)
$N_{prior} = 100$	$\lambda_i \sim N$ (0.91, 0.012)	$\lambda_i \sim N$ (1, 0.012)
	$\gamma_{11} \sim N$ (0.3, 0.013)	$\gamma_{11} \sim N$ (0.24, 0.013)
	$\beta_{21} \sim N$ (0.4, 0.012)	$\beta_{21} \sim N$ (0.35, 0.012)
	$\gamma_{21} \sim N$ (0.36, 0.011)	$\gamma_{21} \sim N$ (0.31, 0.011)
$N_{prior} = 200$	$\lambda_i \sim N$ (0.93, 0.006)	$\lambda_i \sim N$ (1, 0.006)
	$\gamma_{11} \sim N$ (0.3, 0.006)	$\gamma_{11} \sim N$ (0.26, 0.006)
	$\beta_{21} \sim N$ (.4, .006)	$\beta_{21} \sim N$ (0.36, 0.006)
	$\gamma_{21} \sim N$ (0.36, 0.006)	$\gamma_{21} \sim N$ (0.32, 0.006)
	Combination 4 ($\gamma_{11} = 0$, $\beta_{21} = 0.40$, $\gamma_{21} = 0.12$)	
$N_{prior} = 25$	$\lambda_i \sim N$ (1.19, 0.047)	$\lambda_i \sim N$ (1, 0.047)
	$\gamma_{11} \sim N$ (0, 0.052)	$\gamma_{11} \sim N$ (0.11, 0.052)
	$\beta_{21} \sim N$ (0.40, 0.051)	$\beta_{21} \sim N$ (0.51, 0.051)
	$\gamma_{21} \sim N$ (0.12, 0.046)	$\gamma_{21} \sim N$ (0.23, 0.046)
$N_{prior} = 50$	$\lambda_i \sim N$ (1.11, 0.024)	$\lambda_i \sim N$ (1, 0.024)
	$\gamma_{11} \sim N$ (0, 0.026)	$\gamma_{11} \sim N$ (0.08, 0.026)
	$\beta_{21} \sim N$ (0.40, 0.025)	$\beta_{21} \sim N$ (0.48, 0.025)
	$\gamma_{21} \sim N$ (0.12, 0.023)	$\gamma_{21} \sim N$ (0.20, 0.023)
$N_{prior} = 100$	$\lambda_i \sim N$ (1.06, 0.012)	$\lambda_i \sim N$ (1, 0.012)
	$\gamma_{11} \sim N$ (0, 0.013)	$\gamma_{11} \sim N$ (0.06, 0.013)
	$\beta_{21} \sim N$ (0.4, 0.013)	$\beta_{21} \sim N$ (0.46, 0.013)
	$\gamma_{21} \sim N$ (0.12, 0.011)	$\gamma_{21} \sim N$ (0.17, 0.011)
$N_{prior} = 200$	$\lambda_i \sim N$ (1.04, 0.006)	$\lambda_i \sim N$ (1, 0.006)
	$\gamma_{11} \sim N$ (0, 0.006)	$\gamma_{11} \sim N$ (0.04, 0.006)
	$\beta_{21} \sim N$ (0.40, 0.006)	$\beta_{21} \sim N$ (0.44, 0.006)
	$\gamma_{21} \sim N$ (0.12, 0.006)	$\gamma_{21} \sim N$ (0.16, 0.006)
	Combination 5 ($\gamma_{11} = 0.40$, $\beta_{21} = 0$, $\gamma_{21} = 0.12$)	
$N_{prior} = 25$	$\lambda_i \sim N$ (1.19, 0.047)	$\lambda_i \sim N$ (1, 0.047)
	$\gamma_{11} \sim N$ (0.40, 0.051)	$\gamma_{11} \sim N$ (0.51, 0.051)
	$\beta_{21} \sim N$ (0, 0.064)	$\beta_{21} \sim N$ (0.13, 0.064)
	$\gamma_{21} \sim N$ (0.12, 0.065)	$\gamma_{21} \sim N$ (0.25, 0.065)
$N_{prior} = 50$	$\lambda_i \sim N$ (1.11, 0.024)	$\lambda_i \sim N$ (1, 0.024)
	$\gamma_{11} \sim N$ (0.40, 0.026)	$\gamma_{11} \sim N$ (0.48, 0.026)
	$\beta_{21} \sim N$ (0, 0.032)	$\beta_{21} \sim N$ (0.09, 0.032)
	$\gamma_{21} \sim N$ (0.13, 0.032)	$\gamma_{21} \sim N$ (0.21, 0.032)
$N_{prior} = 100$	$\lambda_i \sim N$ (1.06, 0.012)	$\lambda_i \sim N$ (1, 0.012)
	$\gamma_{11} \sim N$ (0.40, 0.013)	$\gamma_{11} \sim N$ (0.46, 0.013)
	$\beta_{21} \sim N$ (0, 0.016)	$\beta_{21} \sim N$ (0.06, 0.016)
	$\gamma_{21} \sim N$ (0.12, 0.016)	$\gamma_{21} \sim N$ (0.18, 0.016)
$N_{prior} = 200$	$\lambda_i \sim N$ (1.04, 0.006)	$\lambda_i \sim N$ (1, 0.006)
	$\gamma_{11} \sim N$ (0.40, 0.006)	$\gamma_{11} \sim N$ (0.44, 0.006)
	$\beta_{21} \sim N$ (0, 0.008)	$\beta_{21} \sim N$ (0.04, 0.008)
	$\gamma_{21} \sim N$ (0.12, 0.008)	$\gamma_{21} \sim N$ (0.16, 0.008)
	Combination 6 ($\gamma_{11} = 0$, $\beta_{21} = 0$, $\gamma_{21} = 0.12$)	
$N_{prior} = 25$	$\lambda_i \sim N$ (1.19, 0.047)	$\lambda_i \sim N$ (1, 0.047)
	$\gamma_{11} \sim N$ (0, 0.052)	$\gamma_{11} \sim N$ (0.11, 0.052)
	$\beta_{21} \sim N$ (0, 0.052)	$\beta_{21} \sim N$ (0.11, 0.052)
	$\gamma_{21} \sim N$ (.12, 0.052)	$\gamma_{21} \sim N$ (0.23, 0.052)
$N_{prior} = 50$	$\lambda_i \sim N$ (1.11, 0.024)	$\lambda_i \sim N$ (1, 0.024)
	$\gamma_{11} \sim N$ (0, 0.026)	$\gamma_{11} \sim N$ (0.08, 0.026)
	$\beta_{21} \sim N$ (0, 0.026)	$\beta_{21} \sim N$ (0.08, 0.026)
	$\gamma_{21} \sim N$ (0.12, 0.026)	$\gamma_{21} \sim N$ (0.20, 0.026)
$N_{prior} = 100$	$\lambda_i \sim N$ (1.06, 0.012)	$\lambda_i \sim N$ (1, 0.012)
	$\gamma_{11} \sim N$ (0, 0.013)	$\gamma_{11} \sim N$ (0.06, 0.013)
	$\beta_{21} \sim N$ (0, 0.013)	$\beta_{21} \sim N$ (0.06, 0.013)
	$\gamma_{21} \sim N$ (0.12, 0.013)	$\gamma_{21} \sim N$ (0.17, 0.013)
$N_{prior} = 200$	$\lambda_i \sim N$ (1.04, 0.006)	$\lambda_i \sim N$ (1, 0.006)
	$\gamma_{11} \sim N$ (0, 0.006)	$\gamma_{11} \sim N$ (0.04, 0.006)
	$\beta_{21} \sim N$ (0, 0.006)	$\beta_{21} \sim N$ (0.04, 0.006)
	$\gamma_{21} \sim N$ (0.12, 0.006)	$\gamma_{21} \sim N$ (0.16, 0.006)

(*Continued*)

Table 3. Continued.

Note. This table contains specifications of informative inaccurate prior distributions for measurement and structural model parameters in Study 2. The first hyperparameter in the normal priors is the mean, and the second hyperparameter is the variance. Measurement error variances, and (residual) variances of latent variables were assigned accurate priors that have the weight of N_{prior}.

standard deviation of this empirical distribution of composite reliability was used as the standard deviation for the informative priors based on N_{prior} = 25, 50, 100, and 200. Once the standard deviation of composite reliability was computed, the mean hyperparameters for inaccurate priors for loadings were created by adding/subtracting 0.5*sd* from the true composite reliability (0.87) and calculating the values of loadings that would produce this value of composite reliability. The inaccuracy of 0.5*sd* was chosen because it is possible to induce at the level of composite reliability without exceeding the upper bound of this quantity (i.e., 1). Larger amounts of inaccuracy (e.g., 2*sd*) could be induced for structural paths, but not for composite reliability. In conditions with inaccurate priors for loadings, the priors for structural paths and variance parameters (i.e., measurement error variances, the variances of the exogenous latent variable, and the residual variances of the endogenous latent variables) remained accurate.

In conditions where both loadings and structural paths had inaccurate priors with 0.5*sd* of bias in the mean hyperparameter, the direction (sign) of the inaccuracy was the same for the measurement model and the structural model. For 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 conditions 4–6 where $\gamma_{11}\beta_{21} = 0$ the bias was positive. This way the simulation study answers questions about changes in power with negatively biased expectations in inaccurate priors for combinations where $\gamma_{11}\beta_{21} > 0$, and changes in Type I error rates with positively biased expectations in inaccurate priors for combinations where $\gamma_{11}\beta_{21} = 0$. Table 3 displays hyperparameters for inaccurate priors for loadings and structural paths with 0.5*sd* of inaccuracy, and Table 4 displays inaccurate priors for structural paths with 1, 2, and 3*sd* of inaccuracy in the mean hyperparameters.

Convergence diagnostics were performed using the PSRF and trace plots using the same criteria outlined in the general methods section.

Dependent variables in the simulation studies

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 the variance and the bias squared of an estimator, $MSE = \text{var} + bias^2$, and RMSE computed as $\sqrt{MSE}$. Relative bias above 10% was considered problematic (Kaplan, 1988). Values of MSE were very small and unhelpful in distinguishing between methods being compared, so they are included in the tables of results but not discussed in the results section.

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 is a measure of whether a method systematically overestimates or underestimates the value of the mediated effect, and it 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. Values of imbalance closer to zero indicated less systematic over- and underestimation.

Results

Study 1: Bayesian methods with diffuse generic and fully conjugate priors versus ML and distribution of the product

This study compared point and interval summaries obtained using Bayesian methods with diffuse generic and fully conjugate priors to ML point estimates and interval estimates obtained using the distribution of the product. The results for point summaries/estimates will be described first, followed by the results for the interval summaries/estimates. In this section, the posterior summaries obtained using diffuse generic priors will have the subscript DG (e.g., the posterior mean obtained using diffuse generic priors is denoted mean_{DG}), and the posterior summaries obtained using fully conjugate priors will have the subscript FC (e.g., the posterior mode obtained using fully conjugate priors is denoted mode_{FC}). We consider the posterior mode, median, and mean, and make recommendations based on their performance using a given set of priors. Complete tables of results for Monte Carlo studies 1–2 are available at https://doi.org/10.6084/m9.figshare.9784403.v1

A comparison of point estimates and summaries of the mediated effect indicated that mean_{DG} had the highest RMSE values, followed by the median_{DG} in parameter combinations with a positive mediated effect, and followed by the ML point estimate in parameter combinations with a zero mediated effect (Figure 2). The differences between point summaries and the ML point estimates became smaller as sample size increased, however, the mean_{DG} had notably higher RMSE than other point summaries and the ML estimate even at $N = 200$ for Combination 4 and at $N = 400$ for Combinations 1, 3, and 4. An inspection of values of efficiency (defined as the standard deviation over 500 replications) for the 7 point estimators showed that the median_{DG}, mean_{DG}, and the ML point estimate emerged as the least efficient point estimators, and the mode_{DG} and mode_{FC} emerged as the point estimators with the highest efficiency (i.e., the lowest standard deviation over replications). However, the mode_{DG} and mode_{FC} also had more than 10% relative bias in Combinations 1–3 where the indirect effect was positive, and the median_{FC} followed closely (Figure 3).

At $N = 50$ and 100, HPD credibility intervals obtained using fully conjugate and diffuse generic

Table 4. Prior specifications for Bayesian methods with accurate priors and inaccurate priors with *1sd*, *2sd*, and *3sd* inaccuracy in the prior expectations for the structural parameters.

	Accurate	Inaccurate 1sd	Inaccurate 2sd	Inaccurate 3sd
	Combination 1 ($\gamma_{11} = 0.60$, $\beta_{21} = 0.20$, $\gamma_{21} = 0.12$)			
$N_{prior} = 25$	$\gamma_{11} \sim N$ (0.60, 0.050)	$\gamma_{11} \sim N$ (0.38, 0.050)	$\gamma_{11} \sim N$ (0.15, 0.050)	$\gamma_{11} \sim N$ (−0.07, 0.050)
	$\beta_{21} \sim N$ (0.20, 0.088)	$\beta_{21} \sim N$ (−0.10, 0.088)	$\beta_{21} \sim N$ (−0.39, 0.088)	$\beta_{21} \sim N$ (−0.70, 0.088)
	$\gamma_{21} \sim N$ (0.12, 0.087)	$\gamma_{21} \sim N$ (−0.18, 0.087)	$\gamma_{21} \sim N$ (−0.47, 0.087)	$\gamma_{21} \sim N$ (−0.77, 0.087)
$N_{prior} = 50$	$\gamma_{11} \sim N$ (0.60, 0.025)	$\gamma_{11} \sim N$ (0.44, 0.025)	$\gamma_{11} \sim N$ (0.28, 0.025)	$\gamma_{11} \sim N$ (0.12, 0.025)
	$\beta_{21} \sim N$ (0.20, 0.044)	$\beta_{21} \sim N$ (−0.01, 0.044)	$\beta_{21} \sim N$ (−0.22, 0.044)	$\beta_{21} \sim N$ (−0.43, 0.044)
	$\gamma_{21} \sim N$ (0.12, 0.044)	$\gamma_{21} \sim N$ (−0.09, 0.044)	$\gamma_{21} \sim N$ (−0.30, 0.044)	$\gamma_{21} \sim N$ (−0.51, 0.044)
$N_{prior} = 100$	$\gamma_{11} \sim N$ (0.60, 0.012)	$\gamma_{11} \sim N$ (0.49, 0.012)	$\gamma_{11} \sim N$ (0.38, 0.012)	$\gamma_{11} \sim N$ (0.26, 0.012)
	$\beta_{21} \sim N$ (0.20, 0.022)	$\beta_{21} \sim N$ (0.05, 0.022)	$\beta_{21} \sim N$ (−0.10, 0.022)	$\beta_{21} \sim N$ (−0.24, 0.022)
	$\gamma_{21} \sim N$ (0.12, 0.022)	$\gamma_{21} \sim N$ (−0.03, 0.022)	$\gamma_{21} \sim N$ (−0.18, 0.022)	$\gamma_{21} \sim N$ (−0.32, 0.022)
$N_{prior} = 200$	$\gamma_{11} \sim N$ (0.60, 0.006)	$\gamma_{11} \sim N$ (0.52, 0.006)	$\gamma_{11} \sim N$ (0.44, 0.006)	$\gamma_{11} \sim N$ (0.36, 0.006)
	$\beta_{21} \sim N$ (0.20, 0.011)	$\beta_{21} \sim N$ (0.09, 0.011)	$\beta_{21} \sim N$ (−0.01, 0.011)	$\beta_{21} \sim N$ (−0.12, 0.011)
	$\gamma_{21} \sim N$ (0.12, 0.011)	$\gamma_{21} \sim N$ (0.02, 0.011)	$\gamma_{21} \sim N$ (−0.09, 0.011)	$\gamma_{21} \sim N$ (−0.19, 0.011)
	Combination 2 ($\gamma_{11} = 0.30$, $\beta_{21} = 0.40$, $\gamma_{21} = 0.12$)			
$N_{prior} = 25$	$\gamma_{11} \sim N$ (0.30, 0.052)	$\gamma_{11} \sim N$ (0.07, 0.052)	$\gamma_{11} \sim N$ (−0.15, 0.052)	$\gamma_{11} \sim N$ (−0.38, 0.052)
	$\beta_{21} \sim N$ (0.40, 0.055)	$\beta_{21} \sim N$ (0.17, 0.055)	$\beta_{21} \sim N$ (−0.07, 0.055)	$\beta_{21} \sim N$ (−0.30, 0.055)
	$\gamma_{21} \sim N$ (0.12, 0.050)	$\gamma_{21} \sim N$ (−0.10, 0.050)	$\gamma_{21} \sim N$ (−0.33, 0.050)	$\gamma_{21} \sim N$ (−0.55, 0.050)
$N_{prior} = 50$	$\gamma_{11} \sim N$ (0.30, 0.026)	$\gamma_{11} \sim N$ (0.14, 0.026)	$\gamma_{11} \sim N$ (−0.02, 0.026)	$\gamma_{11} \sim N$ (−0.18, 0.026)
	$\beta_{21} \sim N$ (0.40, 0.027)	$\beta_{21} \sim N$ (0.23, 0.027)	$\beta_{21} \sim N$ (0.07, 0.027)	$\beta_{21} \sim N$ (−0.10, 0.027)
	$\gamma_{21} \sim N$ (0.12, 0.025)	$\gamma_{21} \sim N$ (−0.04, 0.025)	$\gamma_{21} \sim N$ (−0.20, 0.025)	$\gamma_{21} \sim N$ (−0.35, 0.025)
$N_{prior} = 100$	$\gamma_{11} \sim N$ (0.30, 0.013)	$\gamma_{11} \sim N$ (0.18, 0.013)	$\gamma_{11} \sim N$ (0.07, 0.013)	$\gamma_{11} \sim N$ (−0.04, 0.013)
	$\beta_{21} \sim N$ (0.40, 0.014)	$\beta_{21} \sim N$ (0.28, 0.014)	$\beta_{21} \sim N$ (0.16, 0.014)	$\beta_{21} \sim N$ (0.05, 0.014)
	$\gamma_{21} \sim N$ (0.12, 0.012)	$\gamma_{21} \sim N$ (0.01, 0.012)	$\gamma_{21} \sim N$ (−0.10, 0.012)	$\gamma_{21} \sim N$ (−0.22, 0.012)
$N_{prior} = 200$	$\gamma_{11} \sim N$ (0.30, 0.006)	$\gamma_{11} \sim N$ (0.21, 0.006)	$\gamma_{11} \sim N$ (0.140, v.006)	$\gamma_{11} \sim N$ (0.06, 0.006)
	$\beta_{21} \sim N$ (0.40, 0.007)	$\beta_{21} \sim N$ (0.32, 0.007)	$\beta_{21} \sim N$ (0.23, 0.007)	$\beta_{21} \sim N$ (0.15, 0.007)
	$\gamma_{21} \sim N$ (0.12, 0.006)	$\gamma_{21} \sim N$ (0.04, 0.006)	$\gamma_{21} \sim N$ (−0.04, 0.006)	$\gamma_{21} \sim N$ (−0.12, 0.006)
	Combination 3 ($\gamma_{11} = 0.30$, $\beta_{21} = 0.40$, $\gamma_{21} = 0.36$)			
$N_{prior} = 25$	$\gamma_{11} \sim N$ (0.30, 0.052)	$\gamma_{11} \sim N$ (0.07, 0.052)	$\gamma_{11} \sim N$ (−0.15, 0.052)	$\gamma_{11} \sim N$ (−0.38, 0.052)
	$\beta_{21} \sim N$ (0.40, 0.047)	$\beta_{21} \sim N$ (0.18, 0.047)	$\beta_{21} \sim N$ (−0.03, 0.047)	$\beta_{21} \sim N$ (−0.25, 0.047)
	$\gamma_{21} \sim N$ (0.36, 0.046)	$\gamma_{21} \sim N$ (0.15, 0.046)	$\gamma_{21} \sim N$ (−0.07, 0.046)	$\gamma_{21} \sim N$ (−0.28, 0.046)
$N_{prior} = 50$	$\gamma_{11} \sim N$ (0.30, 0.026)	$\gamma_{11} \sim N$ (0.14, 0.026)	$\gamma_{11} \sim N$ (−0.02, 0.026)	$\gamma_{11} \sim N$ (−0.18, 0.026)
	$\beta_{21} \sim N$ (0.40, 0.023)	$\beta_{21} \sim N$ (0.25, 0.023)	$\beta_{21} \sim N$ (0.09, v.023)	$\beta_{21} \sim N$ (−0.06, 0.023)
	$\gamma_{21} \sim N$ (0.36, 0.023)	$\gamma_{21} \sim N$ (0.21, 0.023)	$\gamma_{21} \sim N$ (0.06, 0.023)	$\gamma_{21} \sim N$ (−0.09, 0.023)
$N_{prior} = 100$	$\gamma_{11} \sim N$ (0.30, 0.013)	$\gamma_{11} \sim N$ (0.19, 0.013)	$\gamma_{11} \sim N$ (0.07, 0.013)	$\gamma_{11} \sim N$ (−0.04, 0.013)
	$\beta_{21} \sim N$ (0.40, 0.012)	$\beta_{21} \sim N$ (0.29, 0.012)	$\beta_{21} \sim N$ (0.18, 0.012)	$\beta_{21} \sim N$ (0.07, 0.012)
	$\gamma_{21} \sim N$ (0.36, 0.011)	$\gamma_{21} \sim N$ (0.25, 0.011)	$\gamma_{21} \sim N$ (0.15, 0.011)	$\gamma_{21} \sim N$ (0.04, 0.011)
$N_{prior} = 200$	$\gamma_{11} \sim N$ (0.30, 0.006)	$\gamma_{11} \sim N$ (0.21, 0.006)	$\gamma_{11} \sim N$ (0.13, 0.006)	$\gamma_{11} \sim N$ (0.06, 0.006)
	$\beta_{21} \sim N$ (0.40, 0.006)	$\beta_{21} \sim N$ (0.32, 0.006)	$\beta_{21} \sim N$ (0.25, 0.006)	$\beta_{21} \sim N$ (0.17, 0.006)
	$\gamma_{21} \sim N$ (0.36, 0.006)	$\gamma_{21} \sim N$ (0.28, 0.006)	$\gamma_{21} \sim N$ (0.21, 0.006)	$\gamma_{21} \sim N$ (0.13, 0.006)
	Combination 4 ($\gamma_{11} = 0$, $\beta_{21} = 0.40$, $\gamma_{21} = 0.12$)			
$N_{prior} = 25$	$\gamma_{11} \sim N$ (0, 0.052)	$\gamma_{11} \sim N$ (0.23, 0.052)	$\gamma_{11} \sim N$ (0.46, 0.052)	$\gamma_{11} \sim N$ (0.69, 0.052)
	$\beta_{21} \sim N$ (0.40, 0.051)	$\beta_{21} \sim N$ (0.63, 0.051)	$\beta_{21} \sim N$ (0.85, 0.051)	$\beta_{21} \sim N$ (1.08, 0.051)
	$\gamma_{21} \sim N$ (0.12, 0.046)	$\gamma_{21} \sim N$ (0.33, 0.046)	$\gamma_{21} \sim N$ (0.55, 0.046)	$\gamma_{21} \sim N$ (0.76, 0.046)
$N_{prior} = 50$	$\gamma_{11} \sim N$ (0, 0.026)	$\gamma_{11} \sim N$ (0.16, 0.026)	$\gamma_{11} \sim N$ (0.32, 0.026)	$\gamma_{11} \sim N$ (0.48, 0.026)
	$\beta_{21} \sim N$ (0.40, 0.025)	$\beta_{21} \sim N$ (0.56, 0.025)	$\beta_{21} \sim N$ (0.72, 0.025)	$\beta_{21} \sim N$ (0.88, 0.025)
	$\gamma_{21} \sim N$ (0.12, 0.023)	$\gamma_{21} \sim N$ (0.27, 0.023)	$\gamma_{21} \sim N$ (0.42, 0.023)	$\gamma_{21} \sim N$ (0.57, 0.023)
$N_{prior} = 100$	$\gamma_{11} \sim N$ (0, 0.013)	$\gamma_{11} \sim N$ (0.11, 0.013)	$\gamma_{11} \sim N$ (0.23, 0.013)	$\gamma_{11} \sim N$ (0.34, 0.013)
	$\beta_{21} \sim N$ (0.40, 0.013)	$\beta_{21} \sim N$ (0.51, 0.013)	$\beta_{21} \sim N$ (.63, .013)	$\beta_{21} \sim N$ (.74, .013)
	$\gamma_{21} \sim N$ (0.12, 0.011)	$\gamma_{21} \sim N$ (0.23, 0.011)	$\gamma_{21} \sim N$ (0.33, 0.011)	$\gamma_{21} \sim N$ (0.44, 0.011)
$N_{prior} = 200$	$\gamma_{11} \sim N$ (0, 0.006)	$\gamma_{11} \sim N$ (0.08, 0.006)	$\gamma_{11} \sim N$ (0.16, 0.006)	$\gamma_{11} \sim N$ (0.24, 0.006)
	$\beta_{21} \sim N$ (0.40, 0.006)	$\beta_{21} \sim N$ (0.48, 0.006)	$\beta_{21} \sim N$ (0.56, 0.006)	$\beta_{21} \sim N$ (0.64, 0.006)
	$\gamma_{21} \sim N$ (0.12, 0.006)	$\gamma_{21} \sim N$ (0.20, 0.006)	$\gamma_{21} \sim N$ (0.27, 0.006)	$\gamma_{21} \sim N$ (0.35, 0.006)
	Combination 5 ($\gamma_{11} = 0.40$, $\beta_{21} = 0$, $\gamma_{21} = 0.12$)			
$N_{prior} = 25$	$\gamma_{11} \sim N$ (0.40, 0.051)	$\gamma_{11} \sim N$ (0.63, 0.051)	$\gamma_{11} \sim N$ (0.85, v.051)	$\gamma_{11} \sim N$ (1.08, 0.051)
	$\beta_{21} \sim N$ (0, 0.064)	$\beta_{21} \sim N$ (0.25, 0.064)	$\beta_{21} \sim N$ (0.51, v.064)	$\beta_{21} \sim N$ (0.76, 0.064)
	$\gamma_{21} \sim N$ (0.12, 0.065)	$\gamma_{21} \sim N$ (0.37, 0.065)	$\gamma_{21} \sim N$ (0.63, 0.065)	$\gamma_{21} \sim N$ (0.88, 0.065)
$N_{prior} = 50$	$\gamma_{11} \sim N$ (0.40, 0.026)	$\gamma_{11} \sim N$ (0.56, 0.026)	$\gamma_{11} \sim N$ (0.72, 0.026)	$\gamma_{11} \sim N$ (0.88, 0.026)
	$\beta_{21} \sim N$ (0, 0.032)	$\beta_{21} \sim N$ (0.18, 0.032)	$\beta_{21} \sim N$ (0.36, 0.032)	$\beta_{21} \sim N$ (0.54, 0.032)
	$\gamma_{21} \sim N$ (0.13, 0.032)	$\gamma_{21} \sim N$ (0.30, 0.032)	$\gamma_{21} \sim N$ (0.48, 0.032)	$\gamma_{21} \sim N$ (0.66, 0.032)
$N_{prior} = 100$	$\gamma_{11} \sim N$ (0.40, 0.013)	$\gamma_{11} \sim N$ (0.51, 0.013)	$\gamma_{11} \sim N$ (0.63, 0.013)	$\gamma_{11} \sim N$ (0.74, v.013)
	$\beta_{21} \sim N$ (0, 0.016)	$\beta_{21} \sim N$ (0.13, 0.016)	$\beta_{21} \sim N$ (0.25, 0.016)	$\beta_{21} \sim N$ (0.38, 0.016)
	$\gamma_{21} \sim N$ (0.12, 0.016)	$\gamma_{21} \sim N$ (0.25, 0.016)	$\gamma_{21} \sim N$ (0.37, 0.016)	$\gamma_{21} \sim N$ (0.50, 0.016)
$N_{prior} = 200$	$\gamma_{11} \sim N$ (0.40, 0.006)	$\gamma_{11} \sim N$ (0.48, 0.006)	$\gamma_{11} \sim N$ (0.56, 0.006)	$\gamma_{11} \sim N$ (0.64, 0.006)
	$\beta_{21} \sim N$ (0, 0.008)	$\beta_{21} \sim N$ (0.09, 0.008)	$\beta_{21} \sim N$ (0.18, 0.008)	$\beta_{21} \sim N$ (0.27, 0.008)
	$\gamma_{21} \sim N$ (0.12, 0.008)	$\gamma_{21} \sim N$ (0.21, 0.008)	$\gamma_{21} \sim N$ (0.30, 0.008)	$\gamma_{21} \sim N$ (0.39, 0.008)
	Combination 6 ($\gamma_{11} = 0$, $\beta_{21} = 0$, $\gamma_{21} = 0.12$)			
$N_{prior} = 25$	$\gamma_{11} \sim N$ (0, 0.052)	$\gamma_{11} \sim N$ (0.23, 0.052)	$\gamma_{11} \sim N$ (0.46, 0.052)	$\gamma_{11} \sim N$ (0.69, 0.052)
	$\beta_{21} \sim N$ (0, 0.052)	$\beta_{21} \sim N$ (0.23, 0.052)	$\beta_{21} \sim N$ (0.45, 0.052)	$\beta_{21} \sim N$ (0.68, 0.052)
	$\gamma_{21} \sim N$ (0.12, 0.052)	$\gamma_{21} \sim N$ (0.35, 0.052)	$\gamma_{21} \sim N$ (0.58, 0.052)	$\gamma_{21} \sim N$ (0.81, 0.052)
$N_{prior} = 50$	$\gamma_{11} \sim N$ (0, 0.026)	$\gamma_{11} \sim N$ (0.16, 0.026)	$\gamma_{11} \sim N$ (0.32, 0.026)	$\gamma_{11} \sim N$ (0.48, 0.026)

(*Continued*)

Table 4. Continued.

	Accurate	Inaccurate 1sd	Inaccurate 2sd	Inaccurate 3sd
	$\beta_{21} \sim N$ (0, 0.026)	$\beta_{21} \sim N$ (0.16, 0.026)	$\beta_{21} \sim N$ (0.32, 0.026)	$\beta_{21} \sim N$ (0.48, 0.026)
	$\gamma_{21} \sim N$ (0.12, 0.026)	$\gamma_{21} \sim N$ (0.28, 0.026)	$\gamma_{21} \sim N$ (0.44, 0.026)	$\gamma_{21} \sim N$ (0.60, 0.026)
$N_{prior} = 100$	$\gamma_{11} \sim N$ (0, 0.013)	$\gamma_{11} \sim N$ (0.11, 0.013)	$\gamma_{11} \sim N$ (0.23, 0.013)	$\gamma_{11} \sim N$ (0.34, 0.013)
	$\beta_{21} \sim N$ (0, 0.013)	$\beta_{21} \sim N$ (0.11, 0.013)	$\beta_{21} \sim N$ (0.23, 0.013)	$\beta_{21} \sim N$ (0.34, 0.013)
	$\gamma_{21} \sim N$ (0.12, 0.013)	$\gamma_{21} \sim N$ (0.23, 0.013)	$\gamma_{21} \sim N$ (0.35, 0.013)	$\gamma_{21} \sim N$ (0.46, 0.013)
$N_{prior} = 200$	$\gamma_{11} \sim N$ (0, 0.006)	$\gamma_{11} \sim N$ (0.08, 0.006)	$\gamma_{11} \sim N$ (0.16, 0.006)	$\gamma_{11} \sim N$ (0.24, 0.006)
	$\beta_{21} \sim N$ (0, 0.006)	$\beta_{21} \sim N$ (0.08, 0.006)	$\beta_{21} \sim N$ (0.16, 0.006)	$\beta_{21} \sim N$ (0.24, 0.006)
	$\gamma_{21} \sim N$ (0.12, 0.006)	$\gamma_{21} \sim N$ (0.20, 0.006)	$\gamma_{21} \sim N$ (0.28, 0.006)	$\gamma_{21} \sim N$ (0.36, 0.006)

Note. This table contains specifications of informative accurate and inaccurate prior distributions for structural paths in Study 2. The first hyperparameter in the normal priors is the mean, and the second hyperparameter is the variance. In these conditions, the loadings, measurement error variances, and (residual) variances of latent variables were assigned accurate priors that have the weight of N_{prior}.

priors had slightly lower power and lower Type I error rates compared to interval estimates using the distribution of the product. At $N = 200$ and 400, the three interval estimators have almost identical power. Power and Type I error rates were higher for credibility intervals obtained using fully conjugate priors than for credibility intervals obtained using diffuse generic priors (Figure 4). The Type I error rates of HPD intervals obtained using fully conjugate priors still remained below the lower bound of Bradley's robustness criterion and were thus not problematic. Coverage for the distribution of the product confidence intervals and Bayesian credibility intervals with diffuse generic and fully conjugate priors was within the bounds of Bradley's (1978) robustness criterion when the mediated effect was nonzero. In combinations with the true mediated effect equal to 0 and at $N = 50$ and 100, HPD intervals had coverage above 0.975. When $\gamma_{11} = \beta_{21} = 0$ all three interval estimators had coverage above 0.975. At $N = 50$ and 100, credibility intervals obtained using diffuse generic priors had higher interval width than credibility intervals obtained using fully conjugate priors and the distribution of the product confidence limits. At $N = 200$ and 400 the differences in interval width of the three estimators became negligible. The three interval estimators had comparable levels of imbalance, and imbalance was close to zero in all parameter combinations and for all sample sizes.

Study 2: Bayesian methods with accurate and inaccurate informative priors versus ML and distribution of the product

The first part of this section describes the statistical properties of *point* summaries of the mediated effect with accurate priors, point summaries with inaccurate priors for loadings and structural paths that carry 0.5*sd* of inaccuracy in the prior expectation for the parameter, and point summaries with inaccurate priors for structural paths that carry 1, 2, and 3*sd* of inaccuracy in the prior expectation. The second part of this section describes the statistical properties of interval summaries of the mediated effect with accurate priors, inaccurate priors for loadings and structural paths with 0.5*sd* of inaccuracy, and inaccurate priors for structural paths with 1, 2, and 3*sd* of inaccuracy in the prior expectation.

Results for point summaries of the mediated effect

Accurate priors. ML point estimates had higher RMSE values than all point summaries obtained using accurate priors with $N_{prior} = 50\%$ and 25%. A closer look at bias and efficiency indicated that higher values of RMSE of the ML point estimates are due to the lower efficiency of ML point estimates relative to Bayesian point summaries in this situation. In fact, the pattern of RMSE values was almost identical to the pattern of standard deviations over replications for the ML point estimates and the posterior mean, median, and mode obtained using accurate priors that carried 50% and 25% of the weight of the likelihood function. A closer look at bias of the ML point estimates and Bayesian point summaries revealed that when the mediated effect was positive and the priors were accurate, the posterior mean and median had slightly lower bias than the ML estimate, whereas the mode had the highest relative bias and was the only point summary to have more than 10% relative bias when the priors carried 50% of the weight of the likelihood function and $N = 50$. The mode remained the most biased point summary at other sample sizes; however, relative bias did not exceed 10% at any other sample size. When the mediated effect was zero, there were almost no differences in bias between the point summaries, and in most combinations of parameter values and sample size all point summaries obtained using accurate priors had lower absolute bias than the ML estimate.

Inaccurate priors (0.5sd). For the sake of brevity, in the following paragraphs, the posterior summaries

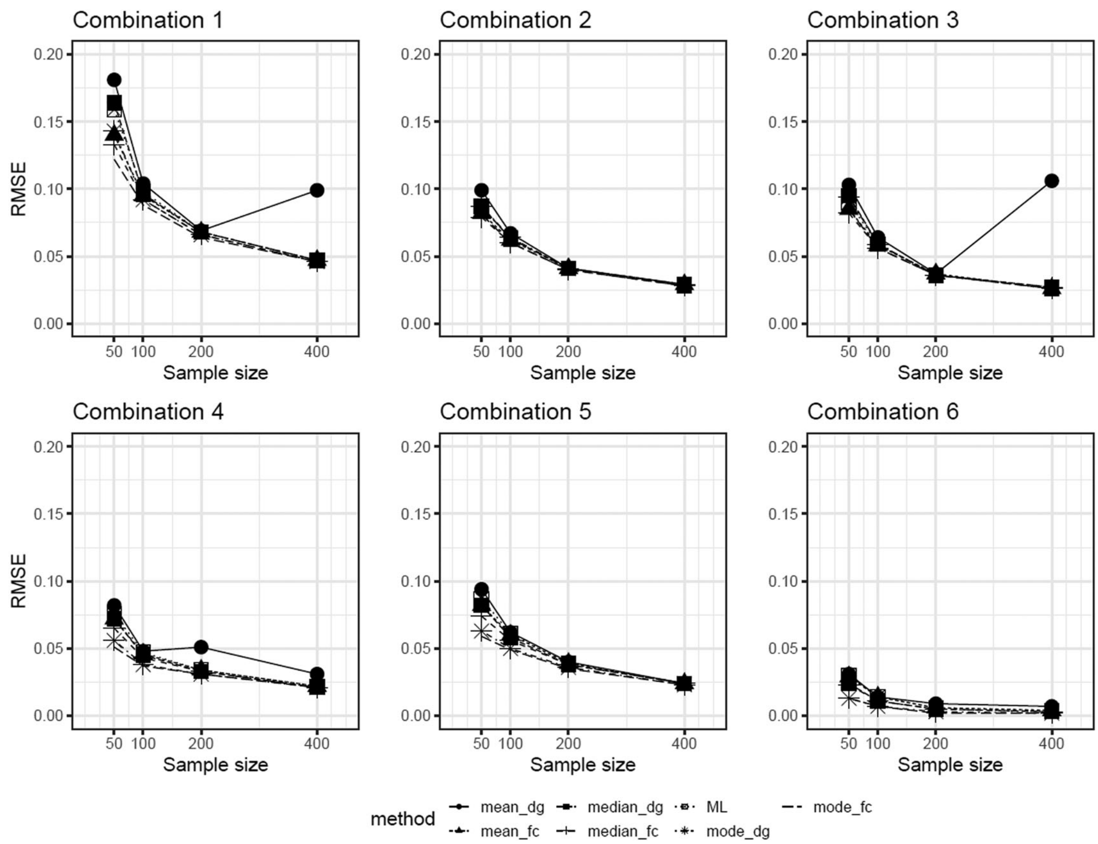

Figure 2. Values of RMSE for ML point estimates and the posterior mean, median, and mode obtained using diffuse generic and fully conjugate priors at $N = 50$, 100, 200, and 400.

with inaccurate priors that carry 0.5*sd* of bias in the prior expectations for structural paths will have a subscript STR (e.g., the posterior mean obtained using inaccurate priors for structural paths will be denoted mean_{STR}), and the posterior summaries with inaccurate priors that carry .5*sd* of bias in the prior expectations for loadings will have a subscript MEAS (e.g., the posterior median with inaccurate priors for the measurement model parameters will be denoted $\text{median}_{\text{MEAS}}$).

A given point summary (mean, median, or mode) obtained using more informative priors (those carrying 50% of the weight of the likelihood function as opposed to only 25%), had lower values of RMSE, thus suggesting that decreasing the weight of the prior led to increases in RMSE. A closer inspection revealed that the pattern of values for efficiency closely mirrored the pattern for RMSE, i.e., higher N_{prior} led to higher efficiency even though the prior expectations were inaccurate. The results for bias indicated that on average, decreasing the weight of the inaccurate prior led to a reduction in bias for a given point summary, but the reduction in bias was not sufficient to "bring" the relative bias of the mode_{STR} below 10% at $N = 50$ and 100. The mean_{STR} had problematic relative bias at $N = 50$ with $N_{\text{prior}} = 50\%$, but not with $N_{\text{prior}} = 25\%$. Also, the $\text{median}_{\text{STR}}$ had problematic relative bias at $N = 50$ and 100 with $N_{\text{prior}} = 50\%$, but the relative bias was no longer above 10% when $N_{\text{prior}} = 25\%$. When the inaccurate priors for loadings carry 50% of the weight of the likelihood function, the relative bias is above 10% for the $\text{mode}_{\text{MEAS}}$, and when inaccurate priors for loadings carry 25% of the weight of the likelihood function, none of the point summaries have absolute relative bias above 10% for any of the sample sizes. Thus, even though inaccurate priors for loadings still lead to suboptimal statistical properties for at least one of the point summaries of the indirect effect, the consequences of the inaccuracy in the expectation of the normal prior can be alleviated by making the prior 2 times less informative. On the other hand, even though reducing the weight of the inaccurate prior for the mean_{STR} and $\text{median}_{\text{STR}}$ yielded levels of relative bias below 10% at all sample sizes, the mode_{STR} still had instances of relative bias above 10% at $N = 50$. Thus, the excessive bias due to inaccurate priors for structural paths can be alleviated by making the prior half as informative at N=100, 200, and 400, but not for all point summaries of the posterior at $N = 50$.

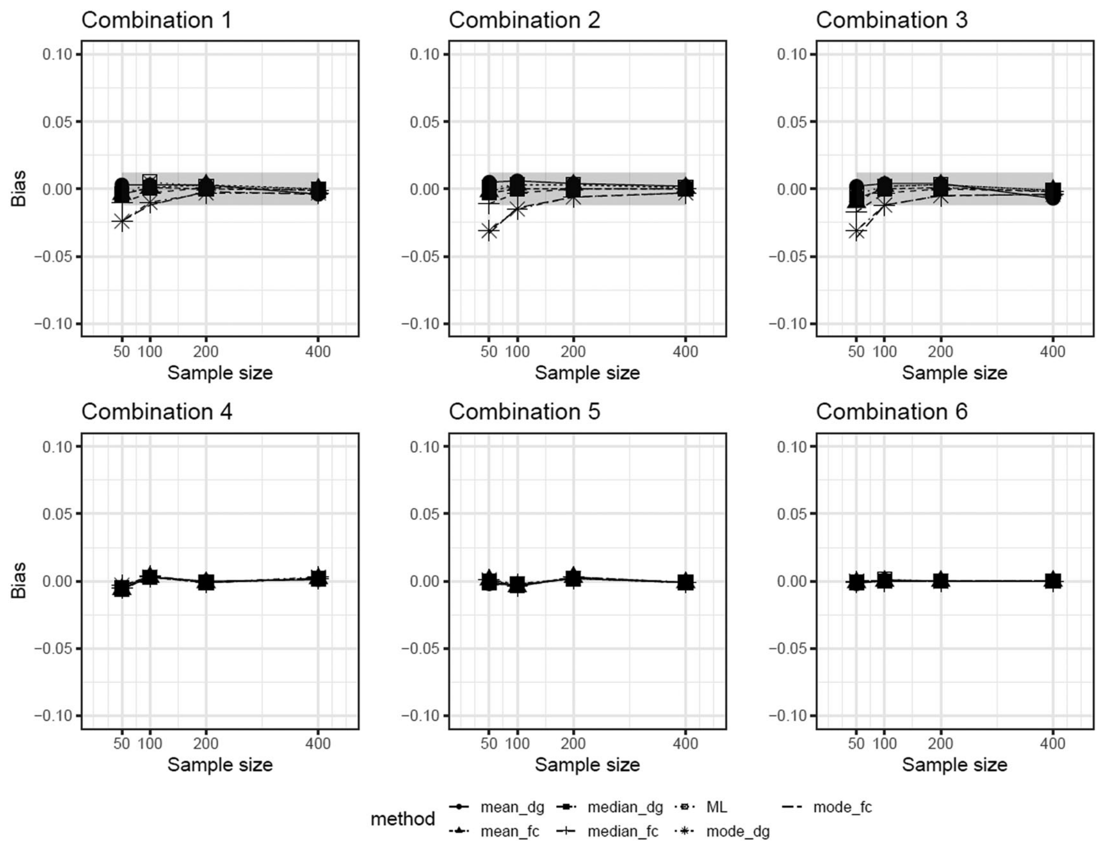

Figure 3. Values of bias for ML point estimates and the posterior mean, median, and mode obtained using diffuse generic and fully conjugate priors at $N = 50$, 100, 200, and 400. The shaded area in plots of combinations 1–3 indicates the area of 10% or less relative bias. Combinations 4–6 do not have values of relative bias because the true mediated effect is 0.

Inaccurate priors (1, 2, 3sd). With inaccurate priors, there was no point summary/estimate that consistently had the lowest RMSE, and the only reoccurring finding in this situation is that the posterior mean had the highest RMSE out of the three point summaries for a given value of prior inaccuracy when the true mediated effect is zero. For a fixed amount of inaccuracy in the prior (1, 2, or 3*sd*), the mean was the least biased point summary and the mode was the most biased point summary when the true mediated effect was positive. Conversely, when the true mediated effect was zero, for a fixed amount of inaccuracy, the mode was the least biased point summary and the mean was the most biased point summary. The posterior mean, median, and mode were more efficient than the ML estimates when the true mediated effect was positive, and had comparable efficiency to Bayesian point summaries with accurate and inaccurate priors when the true mediated effect was zero. More inaccuracy in the prior corresponded with higher efficiency when the true mediated effect was positive and with lower efficiency when the true mediated effect was zero. Reducing the weight of the prior distribution from 50% to 25% of the weight of the likelihood function did not lead to point summaries of the mediated effect with relative bias below 10% at any of the sample sizes. This finding suggests that negative effects of inaccuracy of at least 1*sd* in the mean hyperparameters of priors for structural paths on the bias of point summaries of the mediated cannot be alleviated by reducing the informativeness of the prior by 50%.

Results for interval summaries of the mediated effect

Accurate priors. As expected, in most parameter combinations, 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. The maximum increase in power with accurate priors was 61%, however, accurate priors led to credibility intervals with less power than the distribution of the product confidence limits in parameter Combination 1 where $\gamma_{11} = 0.6$ and $\beta_{21} = 0.2$ at $N = 50$ and 100. Credibility intervals with accurate priors had coverage closer to 1 and more instances of

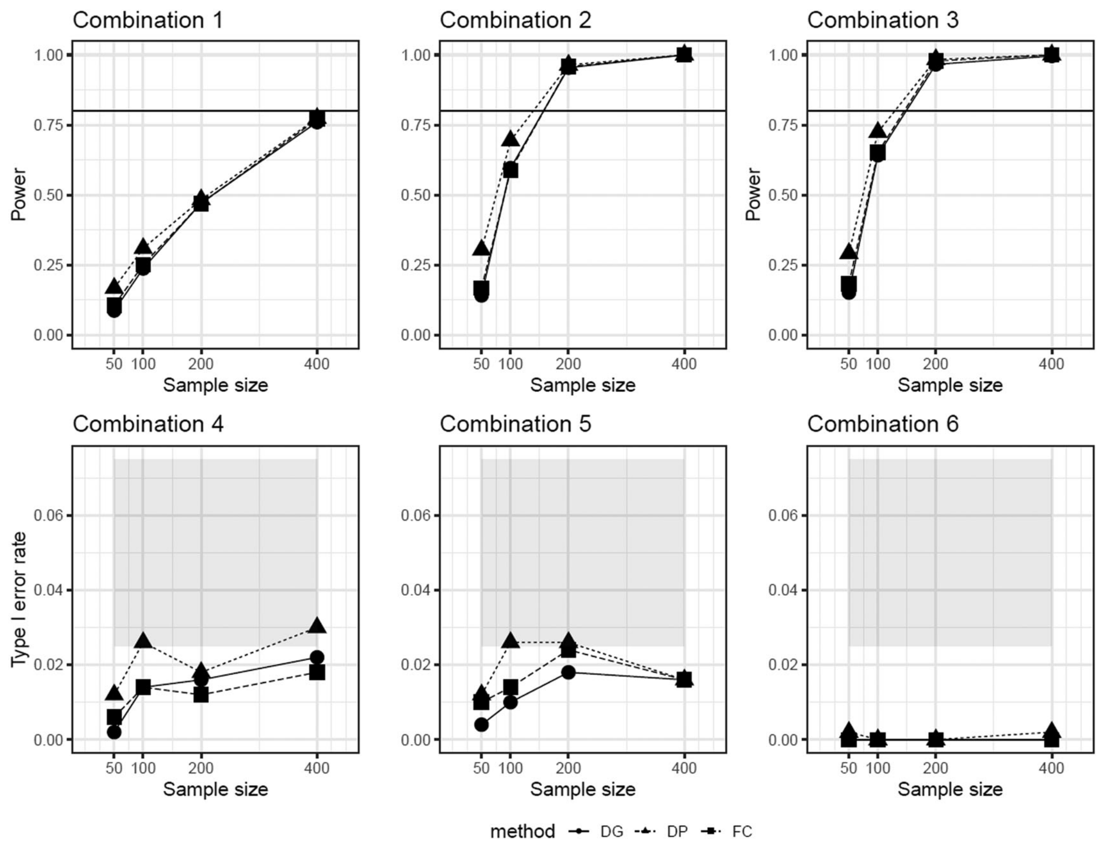

Figure 4. Power and Type I error rates for the distribution of the product confidence limits (DP) and the highest posterior density (HPD) credibility intervals obtained using diffuse generic (DG) and fully conjugate (FC) priors.

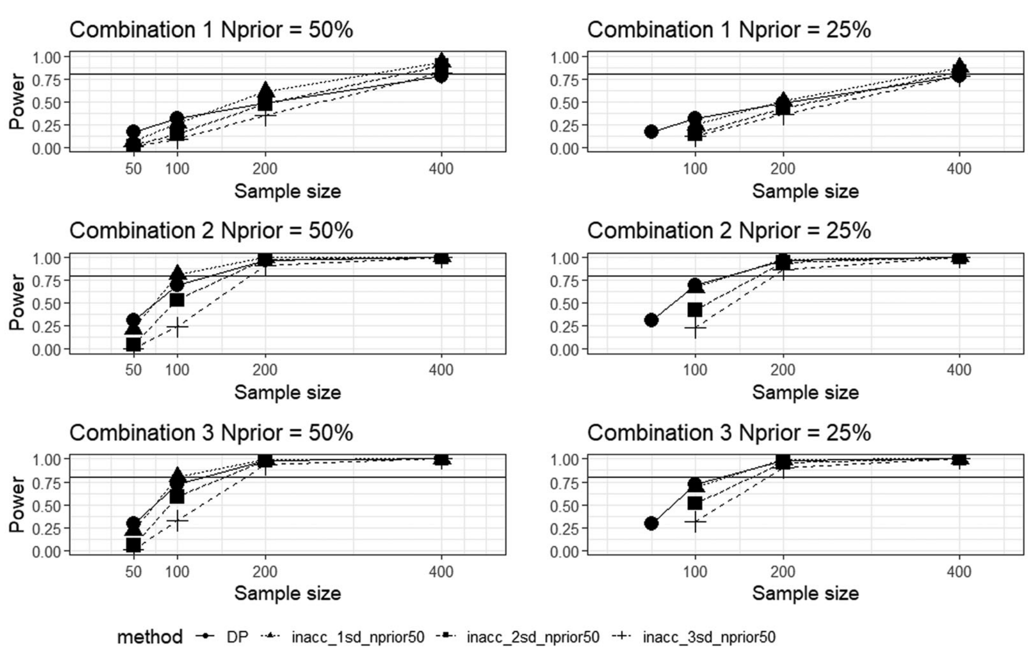

Figure 5. Power for the distribution of the product confidence limits and the highest posterior density (HPD) credibility intervals obtained using informative priors that carry 50% and 25% of the weight of the likelihood function with expectations about structural parameters that were 1*sd* away from the true value (inacc_1sd_nprior50 and inacc_1sd_nprior25), 2*sd* away from the true value (inacc_2sd_nprior50 and inacc_2sd_nprior25), and 3*sd* away from the true value (inacc_3sd_nprior50 and inacc_3sd_nprior25). The horizontal line denotes power of 0.8.

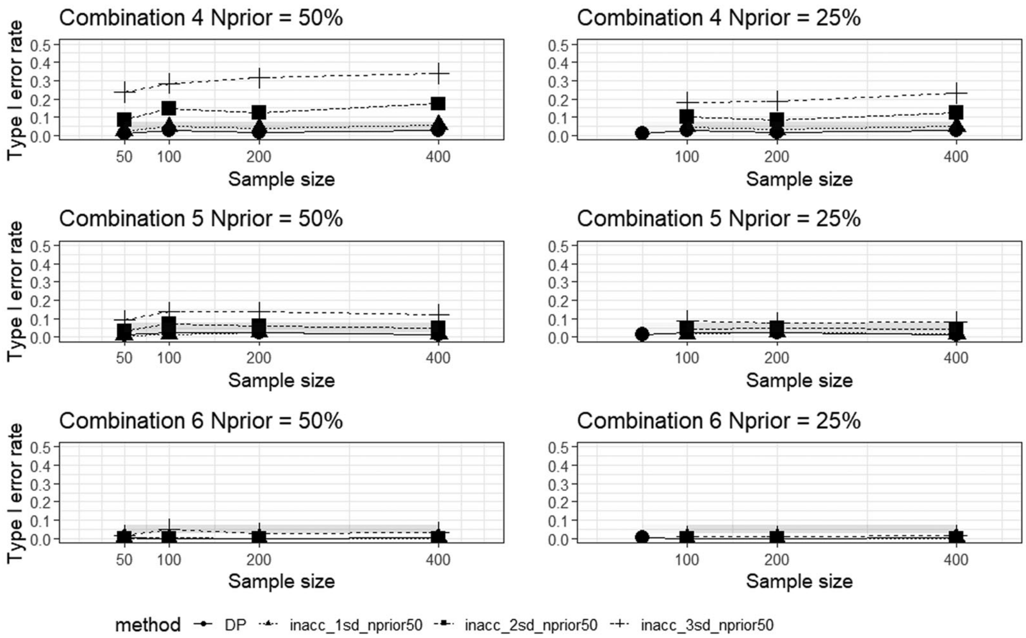

Figure 6. Type I error rates for the distribution of the product confidence limits and the highest posterior density (HPD) credibility intervals obtained using informative priors that carry 50% and 25% of the weight of the likelihood function with expectations about structural parameters that were 1*sd* away from the true value (inacc_1sd_nprior50 and inacc_1sd_nprior25), 2*sd* away from the true value (inacc_2sd_nprior50 and inacc_2sd_nprior25), and 3*sd* away from the true value (inacc_3sd_nprior50 and inacc_3sd_nprior25). The gray band denotes the nominal value of Type I error rates according to Bradley's (1978) robustness criterion.

coverage above 0.975 than distribution of the product confidence limits.

Inaccurate priors (0.5sd). With 0.5*sd* of inaccuracy in the prior expectations, credibility intervals obtained using inaccurate priors for both loadings and structural paths led to power above 0.8 to detect the indirect effect at all sample sizes except at $N = 50$, although at $N = 50$ the power of HPD intervals obtained using inaccurate priors for both loadings and structural paths was equal to or greater than the power of the distribution of the product confidence limits examined in Study 1. Power was lower when the inaccurate priors for both loadings and structural paths carried 25% instead of 50% of the weight of the likelihood function, thus suggesting that even somewhat inaccurate informative priors can lead to increases in power to detect the mediated effect. Also, inaccurate priors with 0.5*sd* of inaccuracy in the prior expectation for loadings and structural paths yielded intervals with lower Type I error rates when the weight of the prior was equal to 50% of the focal sample size than when the weight of the prior was equal to 25% of the size of the focal sample size. Coverage was often above 0.975 with inaccurate priors with 0.5*sd* of inaccuracy; reducing the weight of the inaccurate prior for structural paths from 50% to 25% of the size of the focal sample would often reduce coverage to values within the bounds of Bradley's robustness criterion, whereas coverage above 0.975 remained frequent with inaccurate priors for loadings even when the informativeness of the prior was reduced by half.

Inaccurate priors (1, 2, 3sd). Inaccurate priors in Combinations 1–3 had negatively biased prior expectations, and the maximum decrease in power with inaccurate priors for structural paths was 99%. The decrease in power with the use of inaccurate prior information was larger for parameter Combination 1 (where γ_{11}=0.6 and β_{21}=0.2) than for parameter combinations where γ_{11} and β_{21} were nonzero and comparable in size (Combinations 2 and 3). Also, despite the ratios of weight carried by the prior and the likelihood function being the same at all sample sizes (25% and 50%), at $N = 200$ and 400, inaccurate priors did not consistently decrease power and the decreases in power with higher levels of inaccuracy were notably lower than at $N = 50$ and 100. Also, contrary to the expectations based on theory, at $N = 200$ and 400, even with 1*sd* of inaccuracy in the priors that encoded the assumption that the structural paths were lower than the true values, HPD intervals had more power than the distribution of the product confidence limits in the majority of parameter combinations (Figure 5).

Inaccurate priors for Combinations 4–6 encoded the assumption that the structural parameters are higher than the true values and findings indicate that the resulting Type I error rates exceed 0.075 only with 2*sd* and 3*sd* of inaccuracy in the priors (Figure 6). Coverage was consistently below the lower bound of Bradley's robustness criterion if there was 3*sd* of inaccuracy and in most parameter combinations with 2*sd* of inaccuracy. When the inaccurate priors with 3*sd* of inaccuracy carried half of the weight of the likelihood function, coverage was below 50% in Combinations 2 and 3. Even 1*sd* of inaccuracy tended to produce credibility intervals that had more than twice the absolute value of imbalance of distribution of the product confidence limits. In parameter combinations with a nonzero mediated effect, distribution of the product confidence limits have higher interval width than credibility intervals, and the more inaccurate the prior, the lower the interval width of the credibility intervals. In parameter combinations with a zero mediated effect, higher inaccuracy in the prior was predictive of a higher interval width, however, the distribution of the product still had higher interval width than credibility intervals obtained using inaccurate priors with 1*sd* of inaccuracy.

For a fixed level of inaccuracy, reducing the weight of the prior distribution from 50% to 25% of the weight of the likelihood function led to lower power and lower Type I error rates (but still not within the bounds of Bradley's robustness criterion if the inaccuracy was at least 2*sd* and $\beta_{21} > 0$, and if the inaccuracy was 3*sd* and $\gamma_{11} > 0$). The same reduction in the weight of the prior was sufficient to increase coverage above the lower bound of Bradley's robustness criterion and to notably reduce the imbalance when $\gamma_{11} > \beta_{21} > 0$, but not in other parameter combinations where inaccuracy led to coverage below 0.925. Reducing the weight of informative priors, even inaccurate ones, led to increases in interval width.

Discussion

This project set out to examine the statistical properties of the point and interval estimates/summaries of the mediated effect in the latent variable mediation model obtained using common frequentist methods and Bayesian methods with priors that encode no prior knowledge about the parameters, as well as (accurate and inaccurate) informative priors. Results of Study 1 made it clear that diffuse generic priors yield relatively worse statistical properties for both point and interval summaries and should thus be avoided. These findings are consistent with the previous literature on Bayesian mediation analysis with latent variables and uninformative priors (Chen et al., 2014; van Erp et al., 2018), and with overall findings regarding Bayesian SEM with small samples and uninformative prior distributions (Smid, McNeish, Miočević, & van de Schoot, 2019). ML estimates had less bias and were less efficient than point summaries obtained using fully conjugate priors at smaller sample sizes. Thus if the goal of analysis is to accurately approximate the value of the mediated effect, ML point estimation is preferred over Bayesian methods with both diffuse generic and fully conjugate priors. The decline in statistical properties of interval summaries obtained using Bayesian methods with fully conjugate priors relative to distribution of the product confidence limits is slight. Thus if the probabilistic interpretation of results is important for the research question and there is no relevant prior information available, we recommend researchers use Bayesian methods with diffuse conjugate priors. For example, in the Bayesian framework, it is possible to compute the probability that the indirect effect is greater than a meaningful value (e.g., 0), and that the indirect effect lies within a certain interval (e.g., −0.01 to 0.01; as illustrated by Miočević et al. (2017) for the mediated effect in manifest variable models).

Findings of Study 2 suggest that 0.5*sd* of inaccuracy is more detrimental for approximating the population value of the indirect effect than it is for power, Type I error rates, and coverage of credibility intervals. Furthermore, the excess relative bias can be reduced by making the prior carry half of the weight (by reducing N_{prior} from 50% to 25% of the weight of the likelihood function) at all sample sizes for 0.5*sd* of bias in the prior expectation if the inaccurate priors are assigned to loadings, but not at $N = 50$ if the inaccurate priors are assigned to structural paths. On the other hand, reducing the weight of the prior, even though it is inaccurate, tended to reduce the power, increase Type I error rates (but never above the nominal level of 0.05), and keep coverage above 0.975. When the priors for structural paths had more than 0.5*sd* of bias in the prior expectations, informative priors led to point summaries that were more efficient but also more biased than the ML point estimate. If there was at least 1*sd* of bias in the prior, all point summaries had problematic levels of relative bias that could not be lowered below 10% by reducing the weight of the prior. The comparison of statistical properties of distribution of the product confidence limits and credibility intervals obtained using

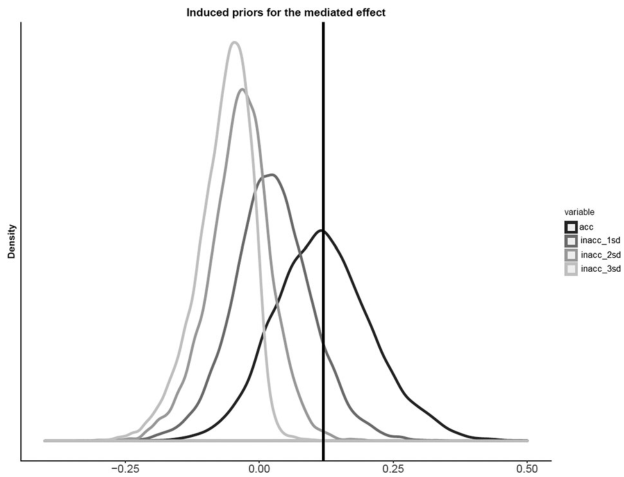

Figure 7. Induced accurate and inaccurate priors for the mediated effect in parameter combination 1 at $N_{\text{prior}} = 100$. The vertical line represents the true value of the indirect effect $\gamma_{11}\beta_{21} = 0.12$. Induced priors on the mediated effect were computed from 1000 simulated draws from accurate and inaccurate priors for paths γ_{11} and β_{21}. The black line represents the accurate prior, the darkest gray line is the inaccurate prior with 1*sd* of inaccuracy in the prior expectation, the next lighter gray line is the inaccurate prior with 2*sd* of inaccuracy in the prior expectation, and the lightest gray line is the inaccurate prior with 3*sd* of inaccuracy in the prior expectation.

inaccurate priors indicated that statistical properties do not always deteriorate with inaccurate priors (e.g., inaccurate priors can yield intervals with more power), however, the risks of using informative priors are not negligible, especially when $N < 200$.

As one of the reviewers pointed out, there is still room for improvement over frequentist methods at any sample size by using Bayesian methods with a prior that is increasingly concentrated at the true value. In the limit, a prior that places the point mass at the true value of a given parameter would maximize accuracy. However, we did not examine these scenarios in the simulation study because it is one that researchers in applied settings are less likely to encounter. It also became apparent that the label "inaccurate prior" ought to be interpreted with respect to the goals of the analysis. The inaccurate priors in this study were inaccurate with respect to the true values of the parameters, but not always with respect to the sign of the true parameter. Thus, in some cases these inaccurate priors created bias in the posterior summaries of the mediated effect, but they did not lead to credibility intervals with lower power. These findings suggest that inaccurate mean hyperparameters in normal priors for structural paths have different consequences depending on the inferential goals of mediation analysis. The goals of the point summaries were to give an accurate representation of the parameter, which was made more difficult by specifying informative priors with mean hyperparameters that are different from the true value of the parameter. The interval summaries were used to conclude that the parameter is not 0 if 0 is outside of the credibility interval. Some of the "inaccurate priors" in Study 2 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. In Study 2 for Combination 1 at $N_{\text{prior}} = 100$, the expectations for γ_{11} and β_{21} with 1*sd* of inaccuracy still induced a prior distribution with the bulk of its density above 0 on the indirect effect (Figure 7). Even though the posterior mean, median, and mode of this distribution were further away from the true value of $\gamma_{11}\beta_{21} = 0.12$ (vertical line in Figure 7), this prior still shifted the posterior density away from 0 and toward positive values, thus yielding higher power. On the other hand, with 2*sd* of inaccuracy, the induced prior to the mediated effect has approximately half of its density below 0, thus no longer favoring positive values of the indirect effect.

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 and power and interval width for interval estimates/summaries, are not as influenced by the accuracy of the priors (at least up to a certain degree) as they are influenced by their informativeness. Imbalance, on the other hand, appears to be more dependent on the (in)accuracy in the hyperparameters for normal priors, and Type I error rates and coverage seem to be influenced by both accuracy and informativeness. It is important to emphasize that all conclusions from the Monte Carlo studies hold only for the parameter values and prior specifications used in this study. Furthermore, the statistical properties of posterior summaries produced with a given inaccuracy in the prior are dependent on the informativeness and size of the focal sample, and the statistical properties of posterior summaries produced with a given informativeness in the prior are dependent on the accuracy and size of the focal sample. 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. When using informative priors, especially for small sample sizes such as $N = 50$ and 100, researchers are advised to put little trust in the point summaries and

Table 5. Recommendations for applied researchers.

1. Are there pros and cons in terms of statistical properties of using Bayesian estimation for the latent variable mediation model in the absence of prior information?
The cons of using Bayesian methods are slightly more bias in the point summaries than the ML point estimate, and slightly less power than the distribution of the product limits at $N = 50$ and 100. The pros of using Bayesian instead of frequentist methods in the absence of prior information are primarily interpretational, i.e., one can compute the probability that the mediated effect lies within a certain range and that it is above/below a certain cutoff value. If the probabilistic interpretation of the findings is important for the research question, it is better to use fully conjugate over diffuse generic priors.
2. What kind of prior distribution should be used when there is no relevant prior information?
In the absence of prior information, fully conjugate priors are preferred over diffuse generic priors, but researchers should be mindful that the informativeness of these priors depends on the scales of the indicators in their study.
3. Are informative priors recommended for Bayesian mediation analysis with latent variables?
At smaller sample sizes ($N = 50$ and 100), informative priors have the potential for causing more damage to statistical properties of the indirect effect if the prior information is inaccurate than for improving the statistical properties of the point and interval summaries of the indirect effect if the prior information is accurate. Researchers should be aware that an informative prior has more potential for yielding inaccurate results at smaller sample sizes, and so should take more care in specifying an informative prior in those cases, possibly also conducting additional analyses (e.g., prior predictive checks and sensitivity analyses). Note that point summaries using informative priors at small sample sizes often have more than 10% relative bias, and if at all possible, we recommend interpreting only the interval summaries of the mediated effect when the sample size is small and abandoning the goal of accurately computing the value of the indirect effect. At more typical sample sizes for SEM ($N = 200$ and 400), informative priors can still cause excessive bias and Type I error rates but are less detrimental to power.

to only use the credibility intervals for making inferences about the mediated effect.

Another notable finding from Study 2 is that even though the ratios of weight allocated to the prior relative to the likelihood function remained the same at all sample sizes (25% and 50%), inaccurate priors had more detrimental effects at $N = 50$ and 100 than at $N = 200$ and 400. The results suggest that using informative priors at sample sizes of 50 and 100 seems to be very risky given that applied researchers have no way of knowing the amount of bias in their prior expectations. Researchers can engage in several activities that may alert them to the possibility of heavily biased prior expectations. First, a researcher may solicit a prior from other experts. If those yield meaningfully different priors that could signal that the researcher's prior may be missing key ideas. Second, a researcher may evaluate the match between the prior and the data *via* prior predictive checks (Box, 1980). Third, researchers could plot the implied plausible parameter space that stems from the set of chosen prior specifications (for instructions and an example of how to do so, see the chapter by Van de Schoot, Veen, Smeets, Winter, & Depaoli, 2020). Fourth, researchers can conduct a sensitivity analysis to examine whether different plausible prior specifications yield similar posterior distributions for the indirect effect; obtaining different results with other plausible prior distributions is a reason to rethink the chosen prior. Fifth, the researcher can ask herself whether the data from the previous study can be considered additional observations from this study, i.e., are the previous and these studies *exchangeable* or do they differ on any relevant characteristics with respect to the research question? The less exchangeable the previous and this study, the higher the risk that the prior distribution based on the previous study will form an inaccurate prior to this study. For one of the first descriptions of exchangeability, see De Finetti (1974), and for more on exchangeability in social sciences research, see the chapter by Miočević, Levy, and Savord (2020). However, none of these approaches unambiguously signals that the prior is highly biased, and in applied settings the researchers have no way of knowing whether the chosen prior or the data set (or neither) provide a more accurate representation of the true effect. We provide a set of recommendations for applied researchers interested in fitting the single mediator model with latent variables as a Bayesian SEM based on the results of the three simulation studies. These recommendations are summarized in Table 5.

Limitations

As with all simulation studies, generalizations beyond the conditions examined may not be warranted. Importantly for our focus, due to the way data were simulated in this study, the findings for the conjugate priors can be generalized to other settings where the latent variables are standardized (i.e., have variances equal to 1), but not to cases where the scale of the reference indicator coupled with the choice of unit loading identification lead to different (residual) variances of the latent variables. We do not discourage researchers who choose to use Bayesian methods and lack knowledge about how to select a variance hyperparameter from using fully conjugate priors, but they should keep in mind that the informativeness of such priors depends on the scales of the indicators and latent variables in their study.

Future directions

Some future directions for this line of research are to examine statistical properties of Bayesian methods for SEM with nonnormal data, building on the work of

Falk (2018), Lai (2018), and Pituch and Stapleton (2008) who compared various frequentist methods in this situation. Future studies should also explore ways of creating accurate informative prior distributions that allow for obtaining the benefits of Bayesian analysis in the single mediator model with latent variables at smaller sample sizes.

Article information

Conflict of interest disclosures: Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

Ethical principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding: This work was supported by Grant No. R37DA09757 from the National Institute on Drug Abuse.

Role of the funders/sponsors: None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Acknowledgments: The authors would like to thank the editor, associate editor, and reviewers for their comments on prior versions of this manuscript. The ideas and opinions expressed herein are those of the authors alone, and endorsement by the authors' institutions or the National Institute on Drug Abuse is not intended and should not be inferred.

ORCID

Milica Miočević http://orcid.org/0000-0001-8487-3666
Roy Levy http://orcid.org/0000-0001-7737-9176
David P. MacKinnon http://orcid.org/0000-0003-0866-6010

References

Aitkin, M., & Aitkin, I. (2005). Bayesian inference for factor scores. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift for Roderick P. McDonald* (pp. 207–222). Mahwah, N.J: Lawrence Erlbaum.

Bollen, K. A., & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods & Research*, *21*(2), 205–229. doi:10.1177/0049124192021002004

Box, G. E. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society. Series A (General)*, *383*, 430. doi:10.2307/2982063

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

Brooks, S. P. (1998). Markov chain Monte Carlo method and its application. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *47*, 69–10. doi:10.1111/1467-9884.00117

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*(4), 434–455. doi:10.2307/1390675

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.

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. 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: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:10.3758/BRM.41.2.425

Coffman, D. L., & Zhong, W. (2012). Assessing mediation using marginal structural models in the presence of confounding and moderation. *Psychological Methods*, *17*(4), 642–664.

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

De Finetti, B. (1974). *Theory of probability* (1st ed. Vol. 1). New York, NY: John Wiley & Sons.

Depaoli, S. (2014). The impact of inaccurate "informative" priors for growth parameters in Bayesian growth mixture modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(2), 239–252.

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

Falk, C. (2018). Are robust standard errors the best approach for interval estimation with nonnormal data in structural equation modeling? *Structural Equation*

Modeling: A Multidisciplinary Journal, *25*(2), 244–266. doi:10.1080/10705511.2017.1367254.

Falk, C., & 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.

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: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*(1), 61–87. doi:10.1080/00273171.2012.640596

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, *1*(3), 515–534. doi:10.1214/06-BA117A

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Boca Raton, FL: CRC press.

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, *15*(4), 309–326. doi:10.1037/a0020761

Jo, B., Stuart, E. A., MacKinnon, D. P., & Vinokur, A. D. (2011). The use of propensity scores in mediation analysis. *Multivariate Behavioral Research*, *46*(3), 425–452. doi:10.1080/00273171.2011.576624

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:10.1207/s15327906mbr2301_4

Kaplan, D., & Depaoli, S. (2012). Bayesian structural equation modeling. In R. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 650–673). New York, NY: Guilford Press.

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:10.1080/01621459.1995.10476592.

Koopman, J., Howe, M., Hollenbeck, J. R., & Sin, H. P. (2015). Small sample mediation testing: Misplaced confidence in bootstrapped confidence intervals. *Journal of Applied Psychology*, *100*(1), 194–202. doi:10.1037/a0036635

Krull, J. L., & MacKinnon, D. P. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review*, *23*(4), 418–444. doi:10.1177/0193841X9902300404

Lai, K. (2018). Estimating standardized SEM parameters given nonnormal data and incorrect model: Methods and comparison. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(4), 600–620. doi:10.1080/10705511.2017.1392248

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). Charlotte, NC: Information Age Publishing.

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

Little, R. J. (2006). Calibrated Bayes: A Bayes/frequentist roadmap. *The American Statistician*, *60*(3), 213–223. doi:10.1198/000313006X117837.

Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(1), 59–72. doi:10.1207/s15328007sem1301_3

Lomnicki, Z. A. (1967). On the distribution of products of random variables. *Journal of the Royal Statistical Society*, *29*, 513–524. doi:10.1111/j.2517-6161.1967.tb00713.x

Lunn, D.J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS - a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, *10*(4), 325–337. doi:10.1023/A:1008929526011.

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:10.1037/a0027131.

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*(3), 384–389.

MacKinnon, D. P., Lockwood, C., 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:10.1037/1082-989X.7.1.83.

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

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, *12*(1), 23–44. doi:10.1037/1082-989X.12.1.23.

Miočević, M. (2019). A Tutorial in Bayesian mediation analysis with latent variables. *Methodology*, *15*(4), 137–146. doi:10.1027/1614-2241/a000177

Miočević, M., Levy, R., & Savord, A. (2020). The role of exchangeability in sequential updating of findings from small studies and the challenges of identifying exchangeable data sets. In R. Van de Schoot & M. Miočević (Eds.), *Small sample size solutions: A guide for applied researchers and practitioners* (pp. 13–29). Abingdon, UK: Routledge.

Miočević, M., MacKinnon, D. P., & Levy, R. (2017). Power in Bayesian mediation analysis for small sample research. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(5), 666–683. doi:10.1080/10705511.2017.1312407

Pituch, K. A., & Stapleton, L. M. (2008). The performance of methods to test upper-level mediation in the presence of nonnormal data. *Multivariate Behavioral Research*, *43*(2), 237–267. doi:10.1080/00273170802034844

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, *6*(1), 7–11.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, *42*(1), 185–227. doi:10.1080/00273170701341316

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

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

Rosseel, Y. (2012). Iavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. doi:10.18637/jss.v048.i02

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, *7*(4), 422–445. doi: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:10.3102/10769986029004461

Smid, S. C., McNeish, D., Miočević, M., & van de Schoot, R. (2019). Bayesian versus frequentist estimation for structural equation models in small sample contexts: A systematic review. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–31. doi:10.1080/10705511.2019.1577140

Springer, M. D., & Thompson, W. E. (1966). The distribution of independent random variables. *SIAM Journal on Applied Mathematics*, *14*(3), 511–526. doi:10.1137/0114046.

Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software*, *12*(3), 1–16.

Tofighi, D., & MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods*, *43*(3), 692–700. doi:10.3758/s13428-011-0076-x

Valente, M. J., Gonzalez, O., Miočević, M., & MacKinnon, D. P. (2016). 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*, *76*(6), 889–911. doi:10.1177/0013164415618992.

Valeri, L., & VanderWeele, T. J. (2013). Mediation analysis allowing for exposure–mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods*, *18*(2), 137–150. doi:10.1037/a0031034

van de Schoot, R., & Depaoli, S. (2014). Bayesian analyses: Where to start and what to report. *European Health Psychologist*, *16*(2), 75–84.

van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & Aken, M. A. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development*, *85*(3), 842–886. doi:10.1111/cdev.12169

Van de Schoot, R., Veen, D., Smeets, L., Winter, S. D., & Depaoli, S. (2020). A tutorial on using the WAMBS-checklist to avoid the misuse of Bayesian statistics. In R. Van de Schoot & M. Miočević (Eds.), *Small sample size solutions: A guide for applied researchers and practitioners* (pp. 30–49). Abingdon, UK: Routledge.

van Erp, S., Mulder, J., & Oberski, D. L. (2018). Prior sensitivity analysis in default Bayesian structural equation modeling. *Psychological Methods*, *23*(2), 363–388. doi:10.1037/met0000162

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