
Structural Equation Modeling: A Multidisciplinary Journal

ISSN: 1070-5511 (Print) 1532-8007 (Online) Journal homepage: https://www.tandfonline.com/loi/hsem20

# Mediation Effects In 2-1-1 Multilevel Model: Evaluation Of Alternative Estimation Methods

Jie Fang, Zhonglin Wen & Kit-Tai Hau

*Structural Equation Modeling: A Multidisciplinary Journal*, 26: 591–606, 2019

ISSN: 1070-5511 print / 1532-8007 online
DOI: https://doi.org/10.1080/10705511.2018.1547967



# Mediation Effects In 2-1-1 Multilevel Model: Evaluation Of Alternative Estimation Methods

Jie Fang,[1] Zhonglin Wen,[2] and Kit-Tai Hau[3]
[1]*Guangdong University of Finance & Economics*
[2]*South China Normal University*
[3]*The Chinese University of Hong Kong*

We compared six common methods in estimating the 2-1-1 (level-2 independent, level-1 mediator, level-1 dependent) multilevel mediation model with a random slope. They were the Bayesian with informative priors, the Bayesian with non-informative priors, the Monte-Carlo, the distribution of the product, the bias-corrected, and the bias-uncorrected parametric percentile residual bootstrap. The Bayesian method with informative priors was superior in relative mean square error (RMSE), power, interval width, and interval imbalance. The prior variance and prior mean were also varied and examined. Decreasing the prior variance increased the power, reduced RMSE and interval width when the prior mean was the true value, but decreasing the prior variance reduced the power when the prior mean was set incorrectly. The influence of misspecification of prior information of the $b$ coefficient on multilevel mediation analysis was greater than that on coefficient $a$. An illustrate example with the Bayesian multilevel mediation was provided.

**Keywords**: Bayesian method, multilevel mediation, prior information

Research in the educational, psychological, medical, managerial, and other disciplines utilizing multilevel data often involves tests of mediation. Importantly, multilevel mediation analyses allow researchers to explore the interactions between constructs at different levels. Specifically, in contrast to the single-level mediation, multilevel mediation analyses help explain how the level-2 (e.g., organization) variables affect the individual-level variables (Fang, Zhang, & Chiou, 2010).

Not surprisingly, therefore, multilevel mediation analyses have attracted great attention resulting in a large number of studies in recent years (Fang et al., 2010; Krull & MacKinnon, 2001; McNeish, 2017; Preacher, Zhang, & Zyphur, 2011; Preacher, Zyphur, & Zhang, 2010; Zhang, Zyphur, & Preacher, 2009). According to the literature review of empirical multilevel mediation studies by McNeish (2017), the most frequently used multilevel mediation model is the 2-1-1 [i.e., 2 (independent at Level 2)–1 (mediator at Level 1)–1 (dependent at Level 1)] mediation model in which the only independent variable $X$ is assessed at level 2 of a two-level model (see Figure 1). The second most frequently used model is the 1-1-1 mediation model in which the independent variable $X$, the mediator $M$, and the dependent variable $Y$ are all measured at level 1 of a two-level hierarchy. The purpose of this research was to compare different methods in the analyses of multilevel mediation and to make recommendations in choosing the most suitable method for the certain specific condition.

## LITERATURE REVIEW

Mediation analysis is a statistical approach used to understand how an independent variable $X$ affects a dependent variable $Y$ through a mediator $M$. For the analysis of simple mediation models, most researchers recommend the use of the bootstrap method, the distribution of the product method, the Monte Carlo method and the Bayesian method

Correspondence should be addressed to Zhonglin Wen, School of Psychology, South China Normal University, Guangzhou, 510631, China. E-mail: wenzl@scnu.edu.cn

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/HSEM.

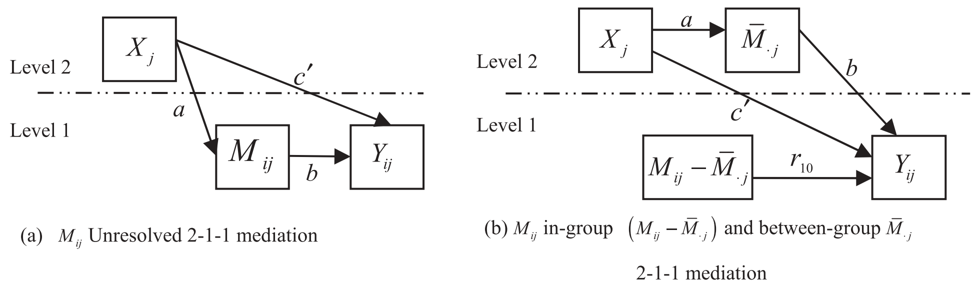


FIGURE 1 Development of the 2-1-1 mediation model. *Note*. 2-1-1 mediation, the three numbers represent the levels of independent variable, the mediation variable and the dependent variable in succession

(see more details below). All these methods analyze the mediating effect by obtaining the asymmetric interval estimates of $\hat{a}\hat{b}$; an interval not including zero implies a significant mediating effect. The advantage this approach is that we do not have to assume the distribution of $\hat{a}\hat{b}$ in the mediation analyses to be normally distributed (Hayes & Scharkow, 2013; MacKinnon, Lockwood, & Williams, 2004; Yuan & MacKinnon, 2009).

Various analytical methods have been compared in previous research. With the 2-1-1 multilevel mediation based on normal data, Pituch, Stapleton, and Kang (2006) compared the performance of several methods including the symmetric interval method, the empirical-M test, the parametric percentile residual bootstrap method, and the bias-corrected parametric percentile residual bootstrap method. The findings showed that the bias-corrected parametric percentile residual bootstrap method performed the best statistically, closely followed by the empirical-M test. Specifically, the bias-corrected parametric percentile residual bootstrap method had the most balanced confidence intervals and the greatest power. The empirical-M test had a power that was similar to that of the bias-corrected parametric percentile residual bootstrap method, but empirical-M test had a slightly worse performance in the balance of confidence intervals. The symmetric interval method had the worst performance among all methods.

Further comparisons on different methods were conducted in two other studies. In Pituch and Stapleton's (2008) simulation study, the performance of the above four methods was further compared in a 2-1-1 multilevel mediation model with non-normal data. The results were consistent with those in the earlier research. In another simulation study, McNeish (2017) compared the performance of the Monte Carlo method and the distribution of the product method in estimating the 2-1-1 multilevel mediation. He found that these two methods had similar power and interval coverage.

The above three comparison studies (McNeish, 2017; Pituch & Stapleton, 2008; Pituch et al., 2006), however, might have several severe limitations that have to be addressed in further studies. First, these comparison studies did not include the Bayesian method, which has attracted increasing research attention in recent years. Simulation studies have shown that the Bayesian method, the bootstrap method and the distribution of the product method had similar performance when used to estimate simple mediation model and moderated mediation models (Biesanz, Falk, & Savalei, 2010; Chen, Choi, Weiss, & Stapleton, 2014; Fang & Zhang, 2012; Miočević, MacKinnon, & Levy, 2017; Wang & Preacher, 2015; Yuan & MacKinnon, 2009), suggesting that the Bayesian method might also be useful for multilevel mediation effect studies. To address such a possibility, therefore, the current study would compare the performance of the Bayesian method (non-informative priors, informative priors), the Monte Carlo method, the distribution of product method, and the parametric residual bootstrap method (bias-uncorrected, bias-corrected) in a 2-1-1 mediation model.

Second, few research guidelines are available to help researchers select prior information in the multilevel mediation analysis. Recently, a simulation study by Miočević et al. (2017) examined how the power in detecting the mediating effect in simple mediation model changed with a gradual increase in the variance of normal prior distribution in the regression coefficient $a$ or $b$. They found that a decrease in the variance led to an increase in power in some sampling conditions. However, their prior mean was set at the true value and the power was the only performance indicator being compared. In our present research, Miočević et al.'s study would be extended to examine the impact of a normal prior to a 2-1-1 random-slope mediation model.

Third, the previous analyses with the 2-1-1 mediation model had serious limitations in the model set up. Specifically, the second part of the mediating effect ($b$, the effect of $M$ on $Y$) was not estimated properly in the mediation model (Pituch & Stapleton, 2008; Pituch et al., 2006; see Figure 1a). This effect could be separated into two parts, one related to the between-group level ($b$), and the other to the within-group level ($r_{10}$) (see Figure 1b)

(Preacher et al., 2011, 2010; Zhang et al., 2009). A new multilevel mediation model (see Figure 1b) is thus necessary to reexamine the methods in Pituch et al.'s mediation analyses.

It is further noted that Pituch et al.'s (2006, 2008) and McNeish's (2017) studies used a 2-1-1 mediation model with a fixed slope, which did not reflect the general multilevel mediation models more commonly used. That is, the 2-1-1 multilevel mediation model with a random slope should be superior to their 2-1-1 multilevel mediation model with a fixed slope, unless the random slope variances so happen to be exactly zero. Specifically, in the current study, we would use the 2-1-1 multilevel mediation model with a random slope instead.

Fourth, in Pituch et al.'s (2006, 2008) and McNeish's (2017) studies, their performance indicator was the confidence interval only, and the point estimate of the mediating effect was not examined. The point estimate is still useful because it is important and easy to understand (e.g., see reports of both point and interval estimates in mediating analyses in Fang & Zhang, 2012; Yuan & MacKinnon, 2009).

In summary, the present research had two major contributions. Firstly, we would make a comprehensive comparison of the six methods discussed above, in which both point and interval estimates would be examined in a 2-1-1 random-slope mediation model. Secondly, we would investigate the impact of the prior distribution in a 2-1-1 random-slope mediation model using several normal prior distributions.

In this article, we would first give a brief introduction of the 2-1-1 multilevel mediation model with a random slope (see Figure 1b) and would then describe the six methods in analyzing multilevel mediation models. We would present the design of the three simulations, followed by their respective results. To help applied researchers in using these methods, we would give an empirical example to illustrate ways of selecting priors in a Bayesian analysis with existing prior information and to facilitate the interpretation of the results so obtained. In the discussion section at the end, we would conclude with recommendations for practical applications of multilevel mediation analysis.

## MULTILEVEL MEDIATION MODEL

In the 2-1-1 multilevel mediation model, a level-2 antecedent influences a level-1 mediator, which in turn affects a Level-1 outcome, as described below,

$$\text{Level 1}: M_{ij} = \beta_{M0j} + \varepsilon_{ij(M)}, \tag{1a}$$

$$\text{Level 2}: \beta_{M0j} = r_{00(M)} + \mathrm{a}X_j + \mu_{0j(M)}, \tag{1b}$$

which represent the effect of the level-2 independent variable $X_j$ on the level-1 mediator variable $M_{ij}$, with subscripts $i$ and $j$ refer to individuals (e.g., employee) and level-2 units (e.g., company), respectively, and $r_{00}$ is the intercept for $M_{ij}$, $a$ is the effect $X_j$ on $M_{ij}$, $\varepsilon_{ij}$ and $\mu_{0j}$ are the level-1 and level-2 residuals, respectively. The remainder of the multilevel mediation model relating to the dependent variable $Y_{ij}$ is represented by the following equations,

$$\text{Level 1}: Y_{ij} = \beta_{Y0j} + \beta_{Y1j}\left(M_{ij} - \overline{M}_{.j}\right) + \varepsilon_{ij(Y)}, \tag{2a}$$

$$\text{Level 2}: \beta_{Y0j} = r_{00(Y)} + c'X_j + b\overline{M}_{.j} + \mu_{0j(Y)}, \text{ and} \tag{2b}$$

$$\text{Level 2}: \beta_{Y1j} = r_{10} + \mu_{1j.} \tag{2c}$$

The coefficient $b$ now indicates the effect of $M_{ij}$ on $Y_{ij}$ at level 2 only, while the coefficient $r_{10}$ represents the effect of $M_{ij}$ on $Y_{ij}$ at level 1 only. As $X_j$ is constant within a given group, variation in $X_j$ cannot influence individual differences within a group. Thus, the effect of $X_j$ could occur at level 2 only, and the mediating effect of $X_j$ on $Y_{ij}$ through $M_{ij}$ could take place at level 2 only. The mediating effect is indicated as $ab$ (Preacher et al., 2010; Zhang et al., 2009). Furthermore, $\mu_{1j}$ is the residual of Level-2 slope and $\beta_{Y1j}$ is a random slope. Specifically, a 2-1-1 multilevel mediation model with a random slope (see Figure 1b and Equations 1 and 2) will be used in the current study.

## METHODS OF MULTILEVEL MEDIATION ANALYSES

### Bayesian method (informative priors, non-informative priors)

The Bayesian method treats parameters as random variables and uses prior information on parameters to obtain the prior distribution of parameters. The posterior distribution of parameters is obtained by combining the observed data with the prior distribution of parameters. In practice, the Markov chain Monte Carlo (MCMC) method is used to approximate the posterior distribution. The mediating effect is computed as the product of the coefficients $a$ and $b$ at each MCMC iteration, so as to obtain the empirical distribution of the mediating effect. The 95% credible interval of the mediating effect is calculated from the values of the 2.5 and 97.5 percentiles in this distribution. The point estimate of the mediating effect can be obtained using the mean of this distribution as well (Fang & Zhang, 2012; Miočević et al., 2017; Yuan & MacKinnon, 2009).

A crucial step in the application of Bayesian method is the selection of the appropriate prior distribution for the model parameters. There are two main categories of priors, which are typically discussed in terms of non-informative and informative priors. Non-informative prior distributions are usually used if no previous information on the parameters is available, and the selected prior distributions should not influence the estimation of the model parameters (Gelman, Carlin, Stern, & Rubin, 2004). An approximately non-informative normal prior can be specified by setting the prior mean to zero and prior variance to $10^3$ or even larger (Miočević et al., 2017). Informative prior distributions incorporate a great deal of the certainty information about the value of the model parameters into the mediation analysis. They can improve the efficiency of estimates when the sample size is small or when the sampling error is large (Depaoli & Clifton, 2015; Yuan & MacKinnon, 2009). Bayesian methods are divided into the Bayesian method with non-informative priors, and the Bayesian method with informative priors.

Sensitivity analysis is helpful in determining how robust the final model results are when priors are modified (van de Schoot, Winter, Ryan, Zondervan-Zwijnenburg, & Depaoli, 2017). This provides a better understanding of the role of the priors in the analysis. Instability of results through sensitivity analysis (e.g., a certain parameter is particularly sensitive to prior settings) will suggest that this prior may be mis-specified and have to be reconsidered. Sensitivity analysis is done after the model has been estimated. It is especially important for small sample research, and it remains a useful practice even with large samples (Miočević et al., 2017; van de Schoot et al., 2017).

## Parametric residual bootstrap method (bias-uncorrected, bias-corrected)

For multilevel models, there are, in principle, two general methods that could be used to resample the data—bootstrapping of cases or bootstrapping of residuals. Generally, bootstrapping of residuals preserves more effectively the structure of multilevel data (Pituch & Stapleton, 2008; Pituch et al., 2006), and for that reason, it would be used in the present study.

The parametric percentile residual bootstrap method which we used in the current study included three steps. First, initial parameter estimates and fitting Equations 1 and 2 were obtained by analyzing the multilevel mediation with the original data set. Second, residual sampling values $\varepsilon^*_{ij}$, $\mu^*_{0j}$ and $\mu^*_{1j}$ were generated from the normal distribution of residuals $\varepsilon_{ij}$, $\mu_{0j}$ and $\mu_{1j}$. Then, $\varepsilon^*_{ij}$, $\mu^*_{0j}$ and $\mu^*_{1j}$ were drawn from the fitting Equations 1 and 2 to get the bootstrap data set $(X_j, M^*_{ij}, Y^*_{ij})$. Parameter estimates $\hat{a}$ and $\hat{b}$ were obtained by analyzing the multilevel mediation with the bootstrap data set. Third, we repeated the earlier step until we had obtained $N$ parameter estimates $\hat{a}$ and $\hat{b}$, which were multiplied to produce an empirical sampling distribution for the mediating effect. The 95% confidence interval would be calculated from the values at 2.5 and 97.5 percentiles in this distribution. The point estimate of the mediating effect could be obtained using the mean of this distribution as well (Pituch & Stapleton, 2008; Pituch et al., 2006).

The bias-corrected parametric percentile residual bootstrap method would make an adjustment to the percentiles used to form the confidence limits based on the degree of bias in the estimation of the mediating effect. A bias arises when the true value of the parameter does not correspond to the median of the distribution of estimates. We made this adjustment by first finding the $z_0$ in the standard normal distribution that corresponded to the percentile position of $\hat{a}_0\hat{b}_0$, the point estimate of the mediating effect calculated in the original data, as relative to the $N$ bootstrap estimates previously obtained. Furthermore, we computed an adjusted critical Z score which was equal to $2Z_0 + Z_{\alpha/2}$ and we identified the percentile in the standard normal distribution associated with this adjusted critical Z score. The upper limit of the confidence interval for the mediating effect was then determined by finding the bootstrap estimate of $\hat{a}\hat{b}$ that corresponded to this percentile. Similarly, for the lower limit of the confidence interval, we computed $2Z_0 - Z_{\alpha/2}$ and determined the bootstrap estimate at this corresponding percentile (2008; Pituch et al., 2006).

## Monte Carlo method

The Monte Carlo (MC) method involved the generation of a sampling distribution $\hat{a}\hat{b}$ using the model estimates and their asymptotic variances and covariance. These estimated values were used to define a multivariate normal (MVN) distribution

$$\begin{bmatrix} a^* \\ b^* \end{bmatrix} \sim MVN\left( \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix}, \begin{bmatrix} \hat{\sigma}^2_a & \hat{\sigma}_{ab} \\ \hat{\sigma}_{ab} & \hat{\sigma}^2_b \end{bmatrix} \right) \tag{3}$$

with $\hat{\sigma}_{ab}$ typically set to 0. Using the parametric assumption in (3), a sampling distribution of $\hat{a}\hat{b}$ was formed by repeatedly generating $a^*$ and $b^*$, which were multiplied to obtain $\hat{a}^*\hat{b}^*$. The 95% confidence interval was then calculated from the values of the 2.5 and 97.5 percentiles in this distribution. The point estimate could be obtained using the mean of this distribution as well. The advantage of the Monte Carlo method is that the raw data are not required and researchers only have to provide the estimates of the mediation paths ($\hat{a}$ and $\hat{b}$) and the standard errors of these paths (MacKinnon et al., 2004; Preacher & Selig, 2012).

### Distribution of the product method

The distribution of the product (DP) method forms a confidence interval for $\hat{a}\hat{b}$ by assuming that the sampling distributions for $\hat{a}$ and $\hat{b}$ are both normal, and the distribution of the product of two normal variables is used (Meeker, Cornwell, & Aroian, 1981). In the current study, the distribution of the product method was obtained using the PRODCLIN program (MacKinnon, Fritz, Williams, & Lockwood, 2007). When the PRODCLIN program is used to estimate the confidence interval of the mediating effect, researchers have to provide the $\hat{a}$ and $\hat{b}$ along with their standard errors only, whereas the raw data are not necessary.

## STUDY 1

The purpose of the first simulation study was to investigate the performance of the above-mentioned six multilevel mediation analysis methods in the estimation of the 2-1-1 multilevel mediation with a random slope. We simultaneously considered the following factors in the simulation study: (a) the level-2 sample size; (b) the level-1 sample size; (c) the mediating effect size, and (d) the analytic method.

First, we systematically manipulated the condition of sample size by varying both the number of clusters and the number of participants within each cluster, with the number of clusters in the study ranging from 10 to 100 (10, 20, 30, 50, 100) and the sample size within clusters in two conditions: 10 and 20. The values selected were similar to those used in previous multilevel simulation studies (Krull & MacKinnon, 2001; McNeish, 2017; Pituch & Stapleton, 2008, Pituch et al., 2006, Pituch, Whittaker, & Stapleton, 2008).

Second, we limited the sizes of mediating effects to three null and three non-null conditions. For the null conditions, $ab = 0$ conditions were obtained by setting (a) $a = b = 0$; (b) $a= 0.39$, $b= 0$; (c) $a= 0$, $b= 0.59$. For the non-null effect size, we used $ab$ effects equal to 0.02 ($a = b = 0.14$), 0.15 ($a = b = 0.39$), and 0.35 ($a = b = 0.59$). These values denoted a small, a medium, and a large mediating effect respectively, as used in previous studies (Cohen, 1988; Fang & Zhang, 2012; MacKinnon et al., 2004; Yuan & MacKinnon, 2009).

In addition, some parameter values in the model were held constant throughout the simulation. Parameter $c'$ from Equation 2 was fixed at the value of 0.1. Parameter $r_{00}$ from Equations 1 and 2 was fixed at the value of 0. Parameter $r_{10}$ from Equation 2 was fixed at the value of 0.14. The residual intraclass correlation was held constant at 0.2 for all conditions. These values were also used in previous multilevel stimulation studies (McNeish, 2017; Pituch et al., 2005, 2006, 2008).

In summary, a 2 (level 1 sample size) × 5 (level 2 sample size) × 6 (size of the mediating effect) factorial design with 60 different conditions was used to evaluate the statistical performance of the aforementioned six different methods in analyzing the 2-1-1 multilevel mediation with a random slope. In each of the conditions, 500 repetitions were computed. The parametric residual bootstrap required 1000 bootstrap samples for each repetition. The MC method required 5000 simulated parameter sets (Hayes & Scharkow, 2013). The distribution of the product method was based on the PRODCLIN program (MacKinnon et al., 2007). The Bayesian method required a total of 1000 burn-in iterations and 10,000 after burn-in iterations so as to generate the Markov chains for the model parameters. The Bayesian method was implemented in WinBUGS. Regression coefficients $a$ and $b$ were assigned normal prior distribution with the prior mean equaled to the true value of $a$ and $b$ respectively, and the prior variance was equal to $10^{-2}$ (Bolin, Finch, & Stenger, 2018; Fang & Zhang, 2012). For all of the other model parameters, non-informative priors were used. More specifically, regression coefficients were assigned normal priors with a mean 0 and a variance $10^6$ (Fang & Zhang, 2012; Yuan & MacKinnon, 2009). Residual variances were assigned inverse gamma priors with the shape and inverse scale equaled to $10^{-3}$ (Fang & Zhang, 2012; Yuan & MacKinnon, 2009).

The simulations were conducted in R (version 3.4.1) and WinBUGS 14. All point estimates of the mediating effect were compared based on the relative mean square error (RMSE), and all interval estimates of the mediating effect were compared on their power, Type I error rate, interval width, and interval imbalance.

## SIMULATION RESULTS

### RMSE

With $MSE_0$ denoting the mean square error (MSE) of the mediating effect estimation on the basis of the parametric percentile residual bootstrap method, the RMSE of an estimate of the mediated effect was defined as $MSE/MSE_0$, in which MSE denoted the MSE of the estimate based on the specific method. The MSE was examined in the equation $E(\hat{a}\hat{b} - ab)^2$, in which $ab$ denoted the true value of the mediating effect, and $\hat{a}\hat{b}$ denoted the point estimate of the mediating effect. The results are shown in Table 1.

The influence of the several factors on the relative mean square error (RMSE) was examined with the analysis of variance (ANOVA). These factors included five types of the level-2 sample size, two types of the level-1 sample size, five types of the method and two types of the mediating effect size (zero or nonzero). The results of ANOVA showed that the interaction of method and level-2 sample

TABLE 1
Relative Mean Square Error (RMSE × 100%) of the Point Estimates of the Mediating Effect in Study 1

| | $a=b=0$ | | $a=0.39, b=0$ | | $a=0, b=0.59$ | | $a=b=0.14$ | | $a=b=0.39$ | | $a=b=0.59$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *N1* | *10* | *20* | *10* | *20* | *10* | *20* | *10* | *20* | *10* | *20* | *10* | *20* |
| | | | | | | N2 = 10 | | | | | | |
| MC | 76.8 | 182.4 | 91.8 | 90.6 | 101.3 | 105.2 | 97.6 | 118.4 | 80.0 | 100.3 | 76.8 | 98.4 |
| DP | 91.6 | 102.5 | 87.8 | 97.0 | 79.5 | 93.5 | 100.6 | 88.6 | 83.9 | 92.9 | 80.0 | 94.5 |
| Boot | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Bayinf | 0.0 | 0.0 | 0.5 | 0.5 | 2.2 | 3.5 | 0.5 | 0.6 | 0.9 | 1.5 | 1.1 | 1.3 |
| Baynon | 72.1 | 157.9 | 109.3 | 104.3 | 91.5 | 105.7 | 129.8 | 87.1 | 77.1 | 85.3 | 88.9 | 99.5 |
| | | | | | | N2 = 20 | | | | | | |
| MC | 106.7 | 109.9 | 88.9 | 85.9 | 99.3 | 111.1 | 96.1 | 90.9 | 103.0 | 125.0 | 119.9 | 97.3 |
| DP | 97.4 | 106.5 | 91.8 | 90.9 | 78.7 | 81.7 | 80.5 | 104.8 | 91.9 | 112.3 | 102.9 | 78.1 |
| Boot | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Bayinf | 0.3 | 0.3 | 2.4 | 2.1 | 9.6 | 14.1 | 2.6 | 3.0 | 5.6 | 5.4 | 4.7 | 4.2 |
| Baynon | 88.8 | 86.8 | 94.7 | 93.5 | 94.9 | 126.6 | 73.2 | 119.0 | 104.0 | 123.1 | 102.4 | 94.5 |
| | | | | | | N2 = 30 | | | | | | |
| MC | 82.8 | 86.1 | 118.3 | 87.0 | 117.5 | 98.7 | 97.5 | 93.9 | 108.9 | 98.1 | 110.0 | 102.0 |
| DP | 105.9 | 124.4 | 96.5 | 99.0 | 101.6 | 104.7 | 88.5 | 80.1 | 108.1 | 100.4 | 115.5 | 73.4 |
| Boot | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Bayinf | 0.9 | 1.2 | 3.9 | 4.3 | 19.1 | 20.5 | 6.0 | 6.6 | 9.2 | 8.9 | 8.8 | 7.1 |
| Baynon | 97.9 | 129.5 | 96.4 | 97.7 | 90.6 | 108.8 | 99.8 | 93.4 | 102.0 | 95.9 | 118.3 | 98.0 |
| | | | | | | N2 = 50 | | | | | | |
| MC | 102.2 | 97.7 | 105.3 | 100.0 | 84.4 | 82.8 | 104.4 | 97.6 | 98.3 | 110.8 | 95.4 | 101.0 |
| DP | 117.1 | 104.7 | 99.9 | 102.2 | 80.8 | 94.8 | 104.5 | 88.2 | 99.3 | 99.6 | 104.1 | 98.9 |
| Boot | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Bayinf | 3.6 | 4.5 | 9.8 | 10.4 | 25.2 | 29.5 | 13.1 | 15.6 | 14.9 | 15.9 | 14.2 | 17.5 |
| Baynon | 98.5 | 134.7 | 101.3 | 110.7 | 85.2 | 86.3 | 85.8 | 91.4 | 106.6 | 95.6 | 108.0 | 104.9 |
| | | | | | | N2 = 100 | | | | | | |
| MC | 79.7 | 108.3 | 107.6 | 98.1 | 77.0 | 90.3 | 100.9 | 92.6 | 94.9 | 99.7 | 108.3 | 91.7 |
| DP | 107.8 | 89.2 | 102.4 | 122.2 | 84.1 | 92.9 | 107.5 | 96.8 | 117.9 | 92.5 | 109.6 | 108.1 |
| Boot | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Bayinf | 8.5 | 17.2 | 26.2 | 26.8 | 41.2 | 56.4 | 29.7 | 29.2 | 33.5 | 31.7 | 35.1 | 30.1 |
| Baynon | 92.4 | 112.1 | 115.2 | 116.9 | 80.7 | 94.2 | 90.4 | 97.2 | 108.2 | 101.4 | 94.3 | 88.7 |

*Note*. Bias-corrected and bias-uncorrected parametric percentile residual Bootstrap method have the same RMSE. MC, DP, Boot, Bayinf, and Baynon refer to the Monte Carlo method, the distribution of the product method, the parametric percentile residual Bootstrap method, the Bayesian method with informative priors, and the Bayesian method with non-informative priors, respectively.

N1 = sample size; N2 = Number of clusters; *a, b* = regression coefficients.

size was statistically significant, $F(16, 200) = 3.0$, $p < .001$, $\eta_p^2 = .19$. Across all other parameters, the RMSEs based on the Bayesian method with informative priors were significantly smaller than the other methods. There was no statistically significant difference among the other methods.

## Power

The power refers to the percentage of replications in which the estimated confidence intervals for the mediating effect did not contain 0 when a true mediating effect existed. Power closer to 1 would be more desirable. The results are shown in Table 2. An ANOVA was conducted that included five types of the level-2 sample size and six types of the method. Results showed that the interaction of the method and the level-2 sample size was not significant, $F(20, 150) = 0.28$, $p = 1.0$, $\eta_p^2 = .04$. The main effect of the method was significant, $F(5, 150) = 2.8$, $p = .02$, $\eta_p^2 = .08$. The Bayesian method with informative priors had the largest power among all the methods. There was no statistically significant difference among the other methods.

## Type I error rate

The Type I error rate referred to the percentage of replications in which the estimated confidence intervals for the mediating effect did not contain 0 when a true mediating effect was zero. The nominal rate of the Type I error was set at 0.05 for all methods. The results are shown in Table 2. Type I error rates outside the Bradley's (1978) robustness criterion range (0.025, 0.075), indicating reasonably close to the 0.05 nominal rate, are indicated in bold numbers in Table 2.

Similar to the analysis with power, an ANOVA was conducted that included the level-2 sample size and the method. Results showed that the interaction of the

TABLE 2
Power, Type I Error Rate, 95% Interval Width and Interval Imbalance for the Mediating Effect in Study 1

| | $a = b = 0$ | | $a = 0.39, b = 0$ | | $a = 0, b = 0.59$ | | $a = b = 0.14$ | | $a = b = 0.39$ | | $a = b = 0.59$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Method* | *Error* | *Width (imban)* | *Error* | *Width (imban)* | *Error* | *Width (imban)* | *Power* | *Width (imban)* | *Power* | *Width (imban)* | *Power* | *Width (imban)* |
| N1 = 10, N2 = 10 | | | | | | | | | | | | |
| MC | **.006** | .526(1) | .030 | .783(−1) | .034 | .748(−3) | .010 | .583(−1) | .086 | .86(3) | .266 | 1.127(19) |
| DP | **.006** | .417(−1) | .038 | .675(−3) | .034 | .625(5) | .014 | .457(6) | .122 | .764(38) | .290 | 1.073(23) |
| Boot | **.002** | .590(1) | .028 | .877(0) | **.020** | .781(−2) | .006 | .646(−1) | .068 | .924(15) | .230 | 1.270(16) |
| BCBoot | **.002** | .631(1) | .056 | .914(4) | .048 | .822(−8) | .012 | .683(1) | .132 | .962(24) | .320 | 1.306(12) |
| Bayinf | **.000** | .043(0) | **.000** | .155(0) | **.000** | .211(0) | .002 | .081(0) | 1.00 | .201(0) | 1.00 | .304(0) |
| Baynon | **.000** | .600(0) | .028 | .883(2) | **.010** | .811(1) | .010 | .645(1) | .050 | .959(11) | .176 | 1.287(17) |
| N1 = 10, N2 = 20 | | | | | | | | | | | | |
| MC | **.002** | .234(−1) | .032 | .438(4) | .024 | .394(−2) | .018 | .277(2) | .266 | .465(20) | .676 | .697(5) |
| DP | **.004** | .180(0) | .048 | .405(−2) | .028 | .361(−2) | .014 | .233(16) | .302 | .445(32) | .664 | .680(9) |
| Boot | **.002** | .243(1) | .030 | .448(5) | .028 | .396(0) | .010 | .301(1) | .200 | .489(16) | .592 | .740(10) |
| BCBoot | **.008** | .260(0) | .056 | .459(−6) | .052 | .410(−2) | .030 | .318(5) | .310 | .507(10) | .662 | .755(4) |
| Bayinf | **.000** | .039(0) | **.002** | .145(−1) | **.002** | .186(−1) | .014 | .076(1) | 1.00 | .187(1) | 1.00 | .281(0) |
| Baynon | **.000** | .239(0) | .042 | .464(−3) | **.022** | .399(1) | .004 | .282(2) | .240 | .498(15) | .608 | .705(17) |
| N1 = 10, N2 = 30 | | | | | | | | | | | | |
| MC | **.002** | .147(−1) | .056 | .319(−4) | .054 | .301(−7) | .026 | .189(1) | .480 | .353(18) | .844 | .534(−1) |
| DP | **.004** | .110(0) | .054 | .310(−1) | .048 | .281(−4) | .024 | .159(0) | .568 | .355(15) | .874 | .531(9) |
| Boot | **.002** | .150(−1) | .060 | .339(−8) | .046 | .294(3) | .020 | .194(−1) | .434 | .374(20) | .822 | .545(4) |
| BCBoot | **.010** | .161(1) | **.086** | .343(−5) | **.078** | .302(9) | .054 | .204(13) | .544 | .385(13) | .860 | .554(−3) |
| Bayinf | **.000** | .036(0) | **.000** | .141(0) | **.004** | .173(0) | .058 | .070(2) | 1.00 | .178(1) | 1.00 | .268(0) |
| Baynon | **.000** | .150(0) | .034 | .232(3) | .036 | .301(−8) | .008 | .191(−2) | .476 | .376(8) | .858 | .547(4) |
| N1 = 10, N2 = 50 | | | | | | | | | | | | |
| MC | **.002** | .084(−1) | .060 | .236(−4) | .058 | .208(9) | .062 | .123(3) | .754 | .270(14) | .964 | .401(5) |
| DP | **.004** | .066(0) | .058 | .229(−7) | .038 | .208(−5) | .072 | .113(35) | .798 | .260(9) | .968 | .394(16) |
| Boot | **.002** | .088(0) | .048 | .244(−8) | **.086** | .213(5) | .052 | .125(2) | .766 | .275(12) | .974 | .407(17) |
| BCBoot | **.004** | .094(1) | .068 | .247(−6) | **.092** | .216(4) | .102 | .131(20) | .802 | .281(5) | .978 | .411(14) |
| Bayinf | **.000** | .031(0) | **.000** | .131(0) | **.014** | .153(−3) | .124 | .064(2) | 1.00 | .160(0) | 1.00 | .244(0) |
| Baynon | **.002** | .088(−1) | .046 | .245(−1) | .054 | .218(−1) | .036 | .126(5) | .750 | .274(10) | .966 | .402(3) |
| N1 = 10, N2 = 100 | | | | | | | | | | | | |
| MC | **.000** | .042(0) | .050 | .160(13) | .040 | .145(2) | .148 | .073(16) | .962 | .184(2) | 1.00 | .274(11) |
| DP | **.006** | .032(1) | .046 | .159(10) | .060 | .143(−4) | .180 | .070(−4) | .946 | .182(8) | 1.00 | .275(−5) |
| Boot | **.002** | .042(−1) | .044 | .163(12) | **.076** | .143(−8) | .168 | .075(12) | .974 | .185(−2) | .998 | .279(0) |
| BCBoot | **.004** | .044(0) | .048 | .163(12) | **.084** | .144(4) | .244 | .078(12) | .976 | .187(−3) | .998 | .280(−3) |
| Bayinf | **.000** | .022(0) | **.004** | .110(2) | **.022** | .121(−1) | .368 | .051(5) | 1.00 | .136(2) | 1.00 | .203(1) |
| Baynon | **.004** | .043(−2) | .058 | .161(14) | .056 | .145(−4) | .146 | .074(23) | .962 | .185(8) | 1.00 | .277(6) |
| N1 = 20, N2 = 10 | | | | | | | | | | | | |
| MC | **.008** | .508(2) | .038 | .751(5) | .028 | .679(−2) | .012 | .573(−1) | .114 | .809(14) | .31 | 1.113(17) |
| DP | **.004** | .394(0) | .054 | .666(1) | .024 | .592(1) | .016 | .455(−2) | .138 | .754(39) | .294 | 1.039(45) |
| Boot | **.002** | .543(1) | **.014** | .887(−3) | **.012** | .691(−2) | .008 | .610(2) | .062 | .913(10) | .212 | 1.285(10) |
| BCBoot | **.002** | .581(1) | .034 | .919(−5) | **.030** | .727(−1) | .016 | .648(3) | .104 | .947(28) | .294 | 1.318(6) |
| Bayinf | **.000** | .042(0) | **.000** | .154(0) | **.000** | .207(0) | .002 | .080(0) | 1.00 | .202(0) | 1.00 | .300(0) |
| Baynon | **.006** | .571(−1) | **.024** | .869(−2) | **.008** | .762(4) | .000 | .627(0) | .052 | .923(12) | .186 | 1.285(18) |
| N1 = 20, N2 = 20 | | | | | | | | | | | | |
| MC | **.002** | .216(1) | .052 | .422(−2) | .032 | .363(2) | .014 | .256(1) | .302 | .464(19) | .61 | .656(18) |
| DP | **.002** | .166(1) | .062 | .398(1) | **.024** | .326(2) | .016 | .220(17) | .302 | .456(20) | .696 | .663(12) |
| Boot | **.000** | .226(0) | .042 | .450(1) | .036 | .368(2) | .014 | .263(2) | .256 | .475(16) | .582 | .702(13) |
| BCBoot | **.002** | .239(−1) | **.078** | .459(−1) | .054 | .379(5) | .022 | .279(−1) | .348 | .490(11) | .640 | .713(3) |
| Bayinf | **.000** | .038(0) | **.000** | .146(0) | **.002** | .185(1) | .006 | .073(2) | 1.00 | .187(0) | 1.00 | .281(−1) |
| Baynon | **.000** | .229(0) | .050 | .443(−5) | .036 | .379(−8) | .014 | .273(−1) | .210 | .479(24) | .626 | .709(10) |
| N1 = 20, N2 = 30 | | | | | | | | | | | | |
| MC | **.000** | .137(0) | .056 | .318(8) | .032 | .271(−8) | .022 | .175(4) | .456 | .348(19) | .842 | .524(5) |
| DP | **.010** | .105(3) | .058 | .314(5) | .048 | .255(−2) | .02 | .154(15) | .526 | .342(21) | .832 | .519(12) |
| Boot | **.000** | .141(0) | .058 | .334(−5) | .054 | .276(−3) | .022 | .185(−1) | .452 | .365(16) | .840 | .539(7) |
| BCBoot | **.008** | .150(−2) | **.092** | .339(−12) | **.088** | .282(−10) | .064 | .195(17) | .548 | .374(7) | .854 | .547(−1) |
| Bayinf | **.000** | .035(0) | **.000** | .139(0) | **.014** | .170(5) | .048 | .070(1) | 1.00 | .176(0) | 1.00 | .263(0) |

(*Continued*)

TABLE 2
(Continued)

| | $a = b = 0$ | | $a = 0.39, b = 0$ | | $a = 0, b = 0.59$ | | $a = b = 0.14$ | | $a = b = 0.39$ | | $a = b = 0.59$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Method* | *Error* | *Width (imban)* | *Error* | *Width (imban)* | *Error* | *Width (imban)* | *Power* | *Width (imban)* | *Power* | *Width (imban)* | *Power* | *Width (imban)* |
| Baynon | **.006** | .143(−1) | .052 | .328(0) | .040 | .284(4) | .020 | .181(1) | .438 | .358(16) | .840 | .540(5) |
| | | | | | N1 = 20, N2 = 50 | | | | | | | |
| MC | **.000** | .080(0) | .050 | .230(7) | .034 | .196(1) | .06 | .118(3) | .746 | .266(9) | .974 | .386(7) |
| DP | **.000** | .060(0) | .042 | .233(−3) | .072 | .193(6) | .088 | .107(36) | .746 | .260(15) | .968 | .386(12) |
| Boot | **.000** | .081(0) | .050 | .239(−1) | .058 | .201(−5) | .048 | .119(9) | .716 | .270(9) | .976 | .397(6) |
| BCBoot | **.004** | 086(0) | .072 | .240(−6) | **.086** | .203(−7) | .100 | .125(32) | .764 | .275(0) | .982 | .400(4) |
| Bayinf | **.000** | .030(0) | **.000** | .129(0) | **.010** | .148(3) | .162 | .060(1) | 1.00 | .158(−1) | 1.00 | .238(1) |
| Baynon | **.002** | .082(1) | .038 | .239(−1) | .040 | .204(−6) | .044 | .117(5) | .712 | .264(5) | .972 | .400(−5) |
| | | | | | N1 = 20, N2 = 100 | | | | | | | |
| MC | **.000** | .039(0) | .056 | .159(2) | .040 | .135(2) | .186 | .072(13) | .964 | .179(−3) | 1.00 | .271(3) |
| DP | **.000** | .028(0) | .058 | .159(−3) | .056 | .133(4) | .204 | .068(35) | .974 | .179(11) | 1.00 | .270(10) |
| Boot | **.002** | .039(1) | .036 | .162(−8) | .074 | .131(−5) | .150 | .072(19) | .960 | .181(14) | 1.00 | .273(−5) |
| BCBoot | **.010** | .041(1) | .042 | .162(−9) | **.096** | .131(−8) | .240 | .074(15) | .966 | .183(6) | 1.00 | .274(−11) |
| Bayinf | **.000** | .022(0) | **.008** | .111(0) | .034 | .114(1) | .382 | .051(2) | 1.00 | .133(2) | 1.00 | .200(−1) |
| Baynon | **.000** | .039(0) | .054 | .161(9) | .056 | .136(0) | .150 | .073(27) | .962 | .181(9) | 1.00 | .272(10) |

*Note*. Values in bold are outside Bradley (1978) robustness criteria. MC, DP, Boot, BCBoot, Bayinf, and Baynon refer to the Monte Carlo method, the distribution of the product method, the parametric percentile residual Bootstrap method, the bias-corrected parametric percentile residual Bootstrap method, the Bayesian method with informative priors, and the Bayesian method with non-informative priors, respectively. Imban denotes the interval imbalance, shown in brackets. Type I error rates outside the Bradley's (1978) robustness criterion range (0.025, 0.075) indicating reasonably close to the 0.05 nominal rate, are indicated in bold numbers.

N1 = sample size; N2 = Number of clusters; *a*, *b* = regression coefficients.

method and the level-2 sample size was not significant, $F(20, 150) = 0.25$, $p = 1.0$, $\eta_p^2 = .03$. The main effect of method was significant, $F(5, 150) = 9.7$, $p < .001$, $\eta_p^2 = .24$. The Bayesian method with informative priors had the lowest rate of the Type I error among all the methods. Fang and Zhang (2012) and Miočević et al. (2017) also suggested that the Bayesian method with informative priors had almost no Type I error regardless of the sample size and the sizes of mediating effects in single-level mediation analyses. The Bias-corrected parametric percentile bootstrap method had the highest rate of Type I error among all the methods. There was no statistically significant difference among the other methods.

## Interval width

The interval width was defined as the difference between the upper confidence limit and the lower confidence limit, with smaller interval width indicating a greater precision of the estimate. The results are shown in Table 2. An ANOVA was conducted that included the level-2 and the level-1 sample size, the method, and whether the mediating effect was zero or nonzero. Results showed that the interaction of the method and the level-2 sample size was significant, $F(20,240) = 5.4$, $p < .001$, $\eta_p^2 = .31$. When the sample sizes of level-2 were 10, 20 and 30, the interval widths of the Bayesian method with informative priors were smaller than all other methods. When the sample size of level-2 was 10, the interval widths of distribution of the product method were smaller than the parametric percentile residual bootstrap method, the bias-corrected parametric percentile residual bootstrap method, and the Bayesian method with non-informative priors.

## Interval imbalance

The interval imbalance was defined as the difference between the numbers of true values that fell on the right side of the interval against those on the left side of it; an imbalance closer to zero was considered more desirable. The results are shown in Table 2. Similar to the analysis with interval width, an ANOVA was conducted. Results showed that the interaction of the method and the level-2 sample size was significant, $F(5, 240) = 8.0$, $p < .001$, $\eta_p^2 = .14$. When the mediating effect size was nonzero, across all other parameters, the interval imbalance of the Bayesian method with informative priors was smaller than that of other methods, and the interval imbalance of the distribution of the product method was greater than that with the other methods. There was no statistically significant difference between the other methods.

In summary, the results of Study 1 suggested that the Bayesian method with an accurate prior information was an excellent way to decrease the RMSE, the interval width, and the imbalance, as well as to increase the power. However, getting an accurate prior information might be difficult in practice. Studies 2 and 3 were designed to examine the influence of misspecification of the normal prior for $a$ or $b$ coefficients in estimating the 2–1-1 multilevel mediation with a random slope.

## STUDY 2

Study 2 examined the influence of the misspecification of prior variance on the regression coefficients $a$ or $b$ in estimating the 2–1-1 multilevel mediation with a random slope using the Bayesian method. As the Bayesian method has unique benefits in small samples (Bolin et al., 2018; McNeish, 2017; Miočević et al., 2017), the number of clusters was set at 10 and 20 and the sample size within clusters was set at 10 in Study 2. Populations with the following values for parameters $a$ and $b$ were simulated: $a = b = 0.14$, 0.39, 0.59, other parameter settings were the same as in Study 1. Random samples were drawn from each of the populations. The point estimation and the interval estimation for the mediating effect were calculated with the different priors for the regression coefficients $a$ and $b$. We fixed the prior mean for regression coefficients $a$ and $b$ as the true value but varied the prior variance. Ten variance conditions were evaluated in the study: five where the prior variance for the coefficient $b$ was set to $10^{-2}$ and the prior variance for the coefficient $a$ varied ($10^{-1}$, $10^{0}$, $10^{1}$, $10^{2}$, $10^{3}$), and five where the prior variance for the coefficient $a$ was set to $10^{-2}$ and the prior variance for the coefficient $b$ varied ($10^{-1}$, $10^{0}$, $10^{1}$, $10^{2}$, $10^{3}$). These values were also used in the Miočević et al. (2017) study.

In summary, a 2 (level 2 sample size) × 3 (size of the mediating effect) × 10 (prior variance conditions) factorial design with 60 different conditions was used to evaluate the statistical performance of misspecification of the prior variance in the analyses of the 2-1-1 multilevel mediation with a random slope. In each condition, 500 repetitions were conducted. The Bayesian method required a total of 1000 burn-in iterations and 10000 after burn-in iterations so as to generate the Markov chains for the model parameters. The simulations were conducted in R (version 3.4.1) and WinBUGS 14. All point estimates of the mediating effect were compared based on their relative mean square error ($MSE_0$ denotes the mean square error of the mediating effect estimation when the prior variance is $10^3$), and all interval estimates of the mediating effect were compared based on their power, the interval width, and the interval imbalance.

## SIMULATION RESULTS

### RMSE

The results are shown in Table 3. The influence of the factors on RMSE was examined with ANOVA. These factors included five types of prior variance, two types of level-2 sample size, and three types of mediating effect size. Results showed no statistically significant interaction but two significant main effects. Increasing the prior variance would increase the RMSE, $F(4, 30) = 74.77$, $p < .001$, $\eta_p^2 = .91$, and increasing sample size would increase the RMSE, $F(1, 30) = 10.33$, $p = .003$, $\eta_p^2 = .26$.

### Power

The results are shown in Table 4. First, similar to analyses with the RMSE, an ANOVA was conducted. Results showed no statistically significant interaction but two significant main effects. Increasing the effect size would increase the power, $F(2, 30) = 82.07$, $p < .001$, $\eta_p^2 = 0.84$, and increasing the sample size would also increase the power, $F(1, 30) = 12.48$, $p = .001$, $\eta_p^2 = .29$.

Second, for small effects ($a = b = 0.14$), there was almost no change in the power as a consequence of increasing the prior variance (Table 4 and Figure 2). Third, for medium and large effects ($a = b = 0.39$ and 0.59), an increase in prior variance led to a smaller power, the reduction in power was relatively slow when the prior variance was greater than $10^0$, but the power was still greater than that of the Bayesian method with non-informative priors (Table 4 and Figure 2). Finally, the

TABLE 3
Relative Mean Square Error (RMSE × 100%) of the Point Estimates of the Mediating Effect in Study 2

| | $a = b = 0.14$ | | $a = b = 0.39$ | | $a = b = 0.59$ | |
|---|---|---|---|---|---|---|
| *prior* | *a = 0.14* | *b = 0.14* | *a = 0.39* | *b = 0.39* | *a = 0.59* | *b = 0.59* |
| | | | N1 = 10, N2 = 10 | | | |
| $10^{-1}$ | 41.1 | 15.9 | 36.8 | 16.6 | 44.9 | 17.8 |
| $10^{0}$ | 86.3 | 61.6 | 69.5 | 68.3 | 75.1 | 69.2 |
| $10^{1}$ | 88.8 | 96.4 | 79.1 | 84.2 | 85.5 | 85 |
| $10^{2}$ | 92.0 | 97.7 | 88.6 | 91.9 | 90.3 | 93.2 |
| $10^{3}$ | 100 | 100 | 100 | 100 | 100 | 100 |
| | | | N1 = 10, N2 = 20 | | | |
| $10^{-1}$ | 56.5 | 35.9 | 63.0 | 38.3 | 74.8 | 33.4 |
| $10^{0}$ | 78.6 | 79.3 | 82.2 | 75.6 | 90 | 86.5 |
| $10^{1}$ | 83.9 | 93.2 | 84.1 | 92.7 | 90.1 | 91.5 |
| $10^{2}$ | 99.9 | 97.9 | 98.7 | 99.9 | 90.8 | 99.9 |
| $10^{3}$ | 100 | 100 | 100 | 100 | 100 | 100 |

N1 = sample size; N2 = Number of clusters; $a$, $b$ = regression coefficients.

TABLE 4
Power, 95% Interval Width and Interval Imbalance for the Mediating Effect in Study 2

| | $a = b = 0.14$ | | | | $a = b = 0.39$ | | | | $a = b = 0.59$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Power | *Width (imban)* | Power | *Width (imban)* | Power | *Width (imban)* | Power | *Width (imban)* | Power | *Width (imban)* | Power | *Width (imban)* |
| Prior | $a$ = 0.14 | | $b$ = 0.14 | | $a$ = 0.39 | | $b$ = 0.39 | | $a$ = 0.59 | | $b$ = 0.59 | |
| | | | | | N1 = 10, N2 = 10 | | | | | | | |
| $10^{-1}$ | 0 | .142(3) | 0 | .177(0) | .628 | .309(9) | .350 | .395(2) | .928 | .466(3) | .808 | .577(1) |
| $10^{0}$ | 0 | .177(2) | 0 | .282(0) | .460 | .387(7) | .240 | .6(−7) | .738 | .577(10) | .420 | .865(2) |
| $10^{1}$ | 0 | .178(8) | 0 | .313(2) | .460 | .400(7) | .210 | .658(4) | .728 | .585(1) | .376 | .981(9) |
| $10^{2}$ | 0 | .184(2) | 0 | .317(3) | .430 | .400(6) | .206 | .692(14) | .700 | .592(6) | .348 | 1.00(4) |
| $10^{3}$ | 0 | .188(3) | 0 | .317(8) | .394 | .408(11) | .198 | .696(3) | .66 | .607(7) | .326 | 1.03(14) |
| | | | | | N1 = 10, N2 = 20 | | | | | | | |
| $10^{-1}$ | .012 | .106(7) | .02 | .140(3) | .874 | .242(3) | .542 | .322(2) | .998 | .356(7) | .926 | .481(3) |
| $10^{0}$ | .008 | .12(10) | .02 | .175(5) | .796 | .266(8) | .39 | .405(0) | .970 | .393(5) | .714 | .596(−1) |
| $10^{1}$ | .008 | .12(8) | .02 | .186(5) | .79 | .266(7) | .366 | .420(2) | .966 | .393(3) | .656 | .617(9) |
| $10^{2}$ | .008 | .12(11) | .02 | .189(4) | .772 | .267(18) | .364 | .428(−2) | .972 | .396(4) | .65 | .627(7) |
| $10^{3}$ | .008 | .12(11) | .02 | .190(9) | .746 | .273(16) | .350 | .430(8) | .972 | .397(4) | .638 | .630(−3) |

N1 = sample size; N2 = Number of clusters; *a, b* = regression coefficients.
Imban denotes the interval imbalance, shown in brackets.

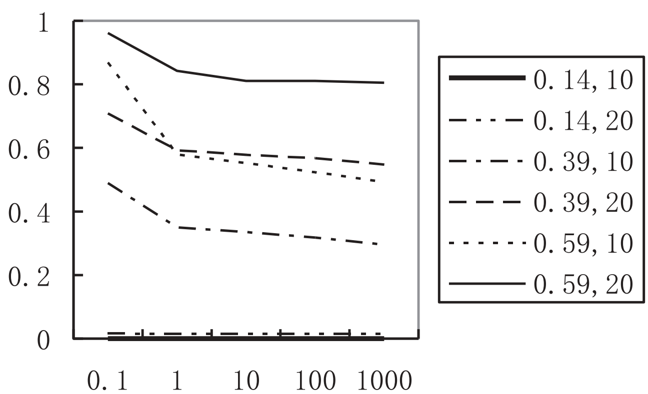


FIGURE 2 Plot of the power as a function of the prior variance for the regression coefficient in study 2. *Note*. (0.14,10) denotes $a = b = 0.14$ and $n_2 = 10$, etc.

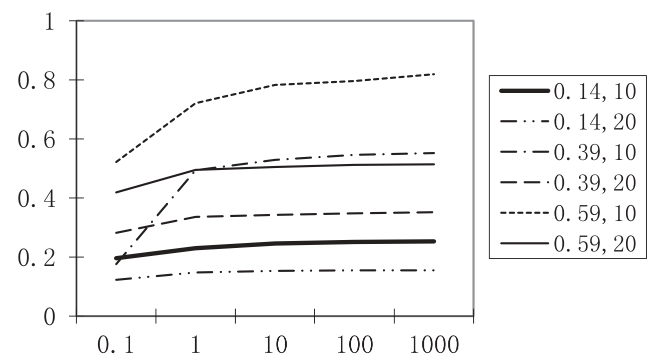


FIGURE 3 Plot of the interval width as a function of the prior variance for the regression coefficient in study 2. *Note*. (0.14,10) denotes $a = b = 0.14$ and $n_2 = 10$, etc.

*t*-test showed that the reduction of the power due to the misspecification of the prior variance of *b* coefficient was significantly greater than that of the coefficient *a*, $t(58) = 2.21$, $p = .031$, $d = 0.6$.

## Interval width and imbalance

The results of interval width are shown in Table 4. First, similar to the analyses with the RMSE, an ANOVA was conducted. Results showed no statistically significant interaction but two significant main effects. Increasing effect size would decrease the interval width, $F(2, 30) = 41.56$, $p < .001$, $\eta_p^2 = .73$, and increasing sample size would decrease the interval width, $F(1, 30) = 18.18$, $p < .001$, $\eta_p^2 = .38$. Second, across all other parameters, an increase in the prior variance would lead to a greater interval width, the increase in interval width was relatively slow when the prior variance was greater than $10^0$, but the interval width was still smaller than that of the Bayesian method with non-informative priors (Table 4 and Figure 3). Third, the *t*-test showed that the increase in interval width due to the misspecification of the prior variance of the *b* coefficient was significantly greater than that of the coefficient *a*, $t(58) = 3.23$, $p = .002$, $d = 0.86$.

The results of interval imbalance are shown in Table 4. Similar to the analyses with the RMSE, an ANOVA was conducted. Results showed that there was no significant effect with the interval imbalance.

# STUDY 3

Study 3 examined the influence of the misspecification of the prior mean on the regression coefficients *a* or *b* in estimating the 2–1-1 multilevel mediation with a random slope using the Bayesian method. As there was almost no change in the power for the small effects ($a = b = 0.14$) in Study 2, populations with the following values for the parameters $a$ and $b$ were simulated: $a = b = 0.39$ and 0.59, other parameter settings were the same as in Study 2. Random samples were drawn from each of the populations. The point estimates and interval estimates for the mediating effect were calculated with the different priors for the regression coefficients $a$ and $b$. We varied both the prior mean and the prior variance. Six mean conditions were evaluated in the study. In one set, the prior mean for the coefficient $b$ was set to the true value and the prior mean for the coefficient $a$ was varied. Specifically, when the true value was 0.59, the prior mean for the coefficient $a$ was set to 0.39 and 0.14; when the true value was 0.39, the prior mean for the coefficient $a$ was set to 0.14. In another set, the prior mean for the coefficient $a$ was set to the true value and the prior mean for the coefficient $b$ was varied. Specifically, when the true value was 0.59, the prior mean for the coefficient $b$ was set to 0.39 and 0.14; when the true value was 0.39, the prior mean for the coefficient $b$ was set to 0.14. Eight variance conditions were evaluated in the study: three where the prior variance for the coefficient $b$ was set to $10^{-2}$ and the prior variance for the coefficient $a$ was varied ($10^{-1}$, $10^{0}$, $10^{1}$, $10^{2}$), and three where the prior variance for the coefficient $a$ was set to $10^{-2}$ and the prior variance for the coefficient $b$ was varied ($10^{-1}$, $10^{0}$, $10^{1}$, $10^{2}$).

In summary, a 2 (level 2 sample size) × 2 (size of the mediating effect) × 6 (prior mean conditions) × 8 (prior variance conditions) factorial design with 192 different conditions was used to evaluate the statistical performance of the misspecification of the normal prior in analyzing the 2-1-1 multilevel mediation with a random slope. In each condition, 500 repetitions were conducted. The Bayesian method required a total of 1000 burn-in iterations and 10,000 after burn-in iterations so as to generate the Markov chains for the model parameters. The simulations were conducted in R (version 3.4.1) and WinBUGS 14. All point estimates of the mediating effect were compared on their relative mean square error ($MSE_0$ denotes the mean square error of the mediating effect estimation when the prior variance is $10^{2}$), and all interval estimates of the mediating effect were compared on their power, the interval width, and the interval imbalance.

## SIMULATION RESULTS

### RMSE

The results are shown in Table 5. The influence of several factors on the RMSE was examined with ANOVA. These factors included four types of prior variance, two types of level-2 sample size, and three types of prior mean. Results showed no statistically significant interactions but two significant main effects. The sample size significantly affected the RMSE, $F(1, 24) = 10.9$, $p = .003$, $\eta_p^2 = .34$, with a larger sample size resulting in a larger RMSE. The prior variance significantly affected the RMSE, $F(4, 24) = 21.25$, $p < .001$, $\eta_p^2 = .73$, with a larger prior variance resulting in a larger RMSE.

TABLE 5
Relative Mean Square Error (RMSE × 100%) of the Point Estimates of the Mediating Effect in Study 3

| | $a = b = 0.39$ | | $a = b = 0.59$ | | $a = b = 0.59$ | |
|---|---|---|---|---|---|---|
| *prior* | *a = 0.14* | *b = 0.14* | *a = 0.14* | *b = 0.14* | *a = 0.39* | *b = 0.39* |
| | | | N1 = 10, N2 = 10 | | | |
| $10^{-1}$ | 62.7 | 29.5 | 81.5 | 50.3 | 51.2 | 60.3 |
| $10^{0}$ | 95 | 66.4 | 87 | 63.6 | 92.2 | 65.3 |
| $10^{1}$ | 99 | 91.1 | 92.4 | 76.9 | 98.5 | 92.1 |
| $10^{2}$ | 100 | 100 | 100 | 100 | 100 | 100 |
| | | | N1 = 10, N2 = 20 | | | |
| $10^{-1}$ | 76.2 | 48.6 | 97.2 | 85.7 | 79.4 | 83.3 |
| $10^{0}$ | 91.4 | 85.9 | 97.2 | 95.2 | 87.8 | 91.7 |
| $10^{1}$ | 98.5 | 90.9 | 99.9 | 97.6 | 97.5 | 98.4 |
| $10^{2}$ | 100 | 100 | 100 | 100 | 100 | 100 |

N1 = sample size; N2 = Number of clusters; *a, b* = regression coefficients.

### Power

The results are shown in Table 6. First, a comparison of the results in Tables 6 and 4 showed that the misspecification of the prior mean for the regression coefficients $a$ or $b$ led to a smaller power. Second, the larger misspecification of prior mean led to a greater reduction in the power when the prior variance was less than $10^{0}$. Specifically, when the true value of the prior mean was 0.59, the power with the 0.14 prior mean was less than that when the prior mean was 0.39 (Table 6). Third, the increase in the prior variance led to the greater power when the prior mean was set to 0.14 (Table 6). Fourth, a larger sample size would result in a higher power. Finally, the $t$-test showed that the reduction of power due to the misspecification of the prior information of the $b$ coefficient was significantly greater than that of the coefficient $a$, $t(46) = 6.45$, $p < .001$.

### Interval width and imbalance

The results of the interval width are shown in Table 6. First, the results showed that the larger sample size resulted in the smaller interval width. Second, the increase in the prior variance led to the greater interval width when the prior mean was misspecified (Table 6). Third, the $t$-test demonstrated that the increase in the interval width due to the misspecification of the prior information of the $b$ coefficient was significantly greater than that of the coefficient $a$, $t(46) = 4.85$, $p < .001$.

TABLE 6
Power, 95% Interval Width and Interval Imbalance for the Mediating Effect in Study 3

| | $a = b = 0.39$ | | | | $a = b = 0.59$ | | | | $a = b = 0.59$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | power | *Width (imban)* | power | *Width (imban)* | power | *Width (imban)* | power | *Width (imban)* | power | *Width (imban)* | power | *Width (imban)* |
| prior | *a*=0.14 | | *b*=0.14 | | *a*=0.14 | | *b*=0.14 | | *a*=0.39 | | *b*=0.39 | |
| | | | | | N1=10, N2=10 | | | | | | | |
| $10^{-1}$ | .368 | .31(15) | .090 | .39(16) | .640 | .465(31) | .210 | .579(53) | .844 | .458(19) | .350 | .868(12) |
| $10^{0}$ | .388 | .384(9) | .172 | .59(13) | .726 | .559(18) | .282 | .902(11) | .736 | .586(5) | .364 | .888(6) |
| $10^{1}$ | .438 | .401(7) | .172 | .685(7) | .730 | .589(1) | .366 | 1.03(-5) | .708 | .604(2) | .354 | .987(5) |
| $10^{2}$ | .456 | .408(7) | .170 | .692(1) | .720 | .594(2) | .372 | 1.07(0) | .712 | .609(2) | .350 | 1.01(11) |
| | | | | | N1=10, N2=20 | | | | | | | |
| $10^{-1}$ | .758 | .24(10) | .282 | .32(19) | .972 | .360(7) | .576 | .481(40) | .982 | .361(-6) | .664 | .611(8) |
| $10^{0}$ | .760 | .27(12) | .336 | .41(17) | .978 | .394(10) | .608 | .598(9) | .980 | .393(7) | .660 | .596(13) |
| $10^{1}$ | .770 | .27(10) | .346 | .425(4) | .976 | .396(7) | .666 | .605(8) | .978 | .400(6) | .650 | .620(10) |
| $10^{2}$ | .770 | .27(13) | .348 | .426(-2) | .970 | .400(9) | .650 | .621(12) | .968 | .401(4) | .666 | .631(1) |

N1 = sample size; N2 = Number of clusters; *a, b* = regression coefficients.
Imban denotes the interval imbalance, shown in brackets.

The results of the interval imbalance are shown in Table 6, however, there was no statistically significant effect on the interval imbalance.

## EMPIRICAL EXAMPLE

In this example, we examined the relationship between team leaders' psychological capital ($X_j$) and team members' organizational citizenship behavior ($Y_{ij}$), with possible mediation through team members' psychological capital ($M_{ij}$). Data were collected from 66 team leaders ($j$ = 66) and 303 team members ($i$ = 303) in a large state-owned enterprise in eastern China (Ren, Wen, Chen, & Ye, 2013). This mediation design was commonly referred to as the 2–1-1 mediation model because the level-2 independent variable $X_j$ was hypothesized to impact the level-1 mediator $M_{ij}$, which in turn was related to the level-1 outcome $Y_{ij}$. The point estimation and intervals estimation for the mediated effects of team leaders' psychological capital ($X_j$) on team members' organizational citizenship behavior ($Y_{ij}$) through team members' psychological capital ($M_{ij}$) were computed using the distribution of the product method, the MC method, the Bayesian method with non-informative priors, and the Bayesian method with informative priors.

The point estimation and the intervals estimation for the mediated effects using the Bayesian methods were computed using *Mplus* (Muthén & Muthén, 1998-2014) and the respective *Mplus* codes to obtain the mediation estimate are provided in the Appendix. The non-informative priors for the Bayesian analysis were selected using the default specification in *Mplus*, with $a \sim N(0, 10^{10})$, $b \sim N(0, 10^{10})$. The regression coefficients ($a$ = 0.41, $b$ = 0.71) and the standard errors ($SE_a$ = 0.004, $SE_b$ = 0.015) from the Ren et al.'s multilevel mediation analysis were set as the mean and standard deviation of the normal priors. In the Ren et al.'s study, the multilevel mediation effect was analyzed under the normal assumption. The RMediation package (Tofighi & MacKinnon, 2011) was used to obtain the interval estimates for the distribution of the product method and the MC method.

None of the intervals constructed using the distribution of the product method, the MC method and the Bayesian methods contained zero, thus one would conclude that the indirect effect of team leaders' psychological capital on team members' organizational citizenship behavior through team members' psychological capital was statistically significant (Table 7). The intervals produced by the Bayesian method with the informative priors were narrower and were thus more precise than those using the Bayesian method with the non-informative priors, the distribution of the product method, and the MC method. These results were consistent with those in our Study 1.

To investigate the influence of the misspecification of the informative priors on the mediation analysis results, we conducted a sensitivity analysis with several values for the prior variance and the prior mean of $a$ and $b$. First, we only varied the prior variance. More specifically, the variance parameter of the priors for the regression coefficients $a$ and $b$ was varied to 5, 10, 20, and 50 times the observed variance of the corresponding regression coefficient. The priors in the sensitivity analyses and the result are summarized in Table 8. These results indicated that the higher the prior variance was, the wider intervals for the mediating effect would become. This observation was consistent with

TABLE 7
Result of the Empirical Example

| *Method* | *Informative prior* | *Mediating effect ab* | *Standard error of ab* | *Interval of ab* |
|---|---|---|---|---|
| DP | \ | 0.260 | 0.060 | [0.151, 0.387] |
| MC | \ | 0.260 | 0.060 | [0.151, 0.388] |
| Baynon | $a \sim N(0, 10^{10})$ $b \sim N(0, 10^{10})$ | 0.256 | 0.064 | [0.143, 0.394] |
| Bayinf | $a \sim N(0.41, 0.004)$ $b \sim N(0.71, 0.015)$ | 0.274 | 0.046 | [0.191, 0.370] |

MC, DP, Bayinf and Baynon refer to the Monte Carlo method, the distribution of the product method, the Bayesian method with informative priors and the Bayesian method with non-informative priors, respectively

*a, b* = regression coefficients

TABLE 8
Result of the Sensitivity Analysis

| | *Informative prior* | *Mediating effect ab* | *Standard error of ab* | *Interval of ab* |
|---|---|---|---|---|
| True value | $a \sim N(0.41, 0.004)$ $b \sim N(0.71, 0.015)$ | 0.274 | 0.046 | [0.191, 0.370] |
| Five times variance | $a \sim N(0.41, 0.02)$ $b \sim N(0.71, 0.075)$ | 0.262 | 0.058 | [0.158, 0.388] |
| Ten times variance | $a \sim N(0.41, 0.04)$ $b \sim N(0.71, 0.15)$ | 0.259 | 0.061 | [0.151, 0.391] |
| Twenty times variance | $a \sim N(0.41, 0.08)$ $b \sim N(0.71, 0.3)$ | 0.258 | 0.062 | [0.147, 0.393] |
| Fifty times variance | $a \sim N(0.41, 0.2)$ $b \sim N(0.71, 0.75)$ | 0.256 | 0.063 | [0.145, 0.394] |
| Half of mean and five times variance | $a \sim N(0.21, 0.02)$ $b \sim N(0.36, 0.075)$ | 0.221 | 0.054 | [0.124, 0.337] |
| Half of mean and fifty times variance | $a \sim N(0.21, 0.2)$ $b \sim N(0.36, 0.75)$ | 0.251 | 0.063 | [0.141, 0.387] |
| Non-informative priors | $a \sim N(0, 10^{10})$ $b \sim N(0, 10^{10})$ | 0.256 | 0.064 | [0.143, 0.394] |

*a, b* = regression coefficients

the results of our second simulation study. Second, we varied both the prior mean and the prior variance. Prior mean was set to half of the observed mean of the corresponding regression coefficient simultaneously as we varied the prior variance. The results are also shown in Table 8. The results indicated that the smaller the prior variance was, the greater the influence the prior mean would have on the mediating effects. This observation was consistent with the results of the third simulation study. Third, the results in Table 8 showed that all the intervals did not contain zero, Robustness of results in a sensitivity analysis indicated that the prior settings were proper.

## DISCUSSION

### Comparisons of methods with informative priors

The first simulation study compared the performance of the Bayesian method (accurate informative priors, non-informative priors), the MC method, the distribution of the product method, and the parametric residual bootstrap method (bias-uncorrected, bias-corrected) in their point and interval estimation of mediating effects in the 2–1-1 multilevel model with a random slope. The results of the point estimation of mediating effects indicated that the RMSEs based on the Bayesian method with informative priors were significantly smaller than those in other methods under all conditions of the sample size. The results on the confidence interval analyses of mediating effects showed that the Bayesian method with informative priors had the largest power, the smallest interval width, and the smallest interval imbalance among the six methods. Those results suggested that the Bayesian method with informative priors might have the best performance in the 2–1-1 random-slope multilevel mediation analysis.

The advantages of the Bayesian method with informative priors in small samples, with the most accurate point and interval estimation, may be closely related to the availability of the prior information. In the situations with small samples, only limited information could be obtained from the data, far from enough for gaining the accurate point and interval estimation of mediating effects. However, when the Bayesian method with informative priors was used, the effective prior information might lead to a shrinkage in the posterior distribution of parameters, thus helping to obtain a more accurate point and interval estimation of mediating effects (Depaoli & Clifton, 2015; Yuan & MacKinnon, 2009). Moreover, the convergence problem caused by negative variances could also be avoided by adding effective prior information (Depaoli & Clifton, 2015).

In larger samples, with more information obtained from the data, the role of prior information became less significant than with smaller samples, and the results of the mediation estimation using different methods became more similar (see Tables 1 and 2). However, the prior information might still have an advantage because with the Bayesian method with informative priors, the RMSEs of the point estimation and the interval width were smaller and the power was larger than when other methods were used. In sum, we recommend the use of the Bayesian method for the multilevel mediation analyses when prior information is available.

## Specifying the prior information

Studies 2 and 3 showed that the misspecification of prior information for the regression coefficients $a$ or $b$ weakened the multilevel mediation analysis. Despite this fact, the result of the multilevel mediation analysis with the misspecification of prior information is generally not worse than that of the Bayesian method with non-informative priors. The larger misspecification of the prior information led to a larger RMSE and a larger interval width, and the change of power, especially when the prior variance is less than 1. Miočević et al. (2017) also suggested that prior variance should be larger than the variances of the parameter estimates obtained from previous studies. If there is no strong theoretical justification for specifying a prior or when a prior was chosen out of convenience, we suggest that a larger than 1 prior variance should be adopted.

It is interesting to note that for medium and large mediating effects ($a = b = 0.39$ and $0.59$), an increase in the prior variance would lead to a smaller power when the prior mean was set to the true value. In contrast, a decrease in the prior variance would lead to a smaller power when the prior mean was not equal to the true value. The possible explanation is that the prior variance reflects researchers' degree of certainty about the prior information. On the one hand, when prior mean is set to the true value, a higher prior variance reflects less confidence that the prior mean is equal to the true value of the regression coefficient $a$ or $b$. This would then create a more diffuse prior for the regression coefficient, which subsequently would lead to a reduction in the power. On the other hand, when the prior mean is not to equal to the true value, a decrease in the prior variance would reflect more confidence in the specified prior mean, and thus would lead to a reduction in the power.

Another notable finding in Studies 2 and 3 was that the influence of the misspecification of the prior information of the $b$ coefficient on the multilevel mediation analysis was significantly greater than that on the coefficient $a$. This result supports the conclusion that it may be more important to focus on the association between the mediator and the outcome than on the association between the intervention and the mediator (Fritz, Taylor, & MacKinnon, 2012). It is noteworthy to point out that the prior information can stem from a meta-analysis, previous studies with comparable research populations, a pilot study, or even experts. Thus, in order to avoid choosing some inappropriate prior information, a sensitivity analysis could be conducted to examine different prior information and to observe the influences of the posterior distribution (Table 8; Depaoli & van de Schoot, 2017; Miočević et al., 2017).

## Comparisons of methods without informative priors

When the prior information is unavailable, the bias-corrected parametric percentile residual bootstrap method was found to have a larger power than other methods in most conditions (see Table 2). However, this method also had higher rates of the Type I error in some situations, suggesting that when this method was used, the decrease in the rates of the Type II error was at the cost of an increase in Type I errors (Fritz et al., 2012). Fritz et al. (2012) clearly stated that when different methods had similar power, it was not necessary to choose a method that would enlarge power at the cost of an increase of the Type I error rate. Moreover, there was also evidence that in the simple mediation analyses on latent and observable variables, the coverage rates of confidence intervals based on the bias-corrected bootstrap method were lower than those based on the bias-uncorrected bootstrap method (Biesanz et al., 2010; Falk & Biesanz, 2015). We, therefore, do not recommend the use of the bias-corrected parametric percentile residual bootstrap method.

Our recommendation was contradictory to that by Pituch et al. (2006, 2008). They argued that the bias-corrected parametric percentile residual bootstrap method performed better than the parametric percentile residual bootstrap method. Worth noting is that their conclusion was based on their finding of an overall (averaging across conditions) Type I error rate of the bias-corrected parametric percentile residual bootstrap method being close to the nominal level. Research has clearly pointed out that the overall error rate is not a suitable indicator of Type I error rate (Biesanz et al., 2010; Falk & Biesanz, 2015).

Furthermore, the distribution of the product method performed the worst in terms of interval imbalance, which was consistent with findings in previous studies (Pituch & Stapleton, 2008, Pituch et al., 2006; Preacher & Selig, 2012). Although McNeish (2017) argued that the Monte Carlo method and the distribution of the product method had similar performance, the interval imbalance was not used to compare the performance of the Monte Carlo method and the distribution of the product method in McNeish's (2017) study.

Among the other three methods (i.e., the Bayesian method with non-informative priors, the MC method, and the parametric percentile residual bootstrap method), which had similar performance, we recommend the use of the MC method among these three for multilevel mediation analyses. The reasons are as follows. First, the MC method can be conducted without the use of the raw data, whereas the bootstrap and the Bayesian method cannot (McNeish, 2017; Preacher & Selig, 2012; Tofighi & MacKinnon, 2016). Second, the MC method is faster to execute than the bootstrap method because a model is fit to the data once only, where a model has to be fitted to all the bootstrap samples in the bootstrap method (Preacher & Selig, 2012). Third, the MC method can be used to analyze multilevel mediation with latent variables, whereas so far the bootstrap method cannot do it (McNeish, 2017, Preacher et al., 2011, 2010). Fourth, the MC method is easy to implement with RMediation (Tofighi & MacKinnon, 2011), whereas there has not been any convenient software for the parametric percentile residual bootstrap method (Preacher et al., 2010). It is also noted that the Bayesian method with non-informative priors could be a useful alternative to the MC method. It is because this method also has a fast computing speed and the ability to analyze multilevel mediation with latent variables (McNeish, 2017).

### Limitations and prospects

First, only four factors (i.e., the sample sizes of levels 1 and 2, the sizes of the mediated effect, and the analytic methods) have been considered in this research. Future research might consider the impacts of other factors, such as ICC (Depaoli & Clifton, 2015; Krull & MacKinnon, 2001), 1-1-1 multilevel models (Bauer, Preacher, & Gil, 2006; Pituch et al., 2008; Yuan & MacKinnon, 2009), and non-normal data (Biesanz et al., 2010; Pituch & Stapleton, 2008).

Second, all the variables in this study were observable variables. Previous studies indicated that there might be a deviation, however, when estimating the mediation with observable variables when the measurement errors have not been considered. Setting up a latent variable model may provide more accurate estimates of the mediating effects (Preacher et al., 2011, 2010). But it is also worth noting McNeish's (2017) found that when the sample size of level 2 was less than or equal to 50 and the MC method and the distribution of the product method were used for the 2–1-1 multilevel mediation analyses, the latent variable approach performed significantly worse than that with observable variables in terms of their power and interval coverage rates. McNeish (2017) further recommended the use of observable variables for multilevel mediation analyses when both observed and latent variables were available and when the sample size of level 2 was less than or equal to 50. These inconsistent findings indicate that the performance of various methods in multilevel mediation analyses has to be further studied.

Third, the present study did not consider multiple mediation analyses. Tofighi and MacKinnon (2016) compared the performance of the MC method and the bootstrap method in SEM analyses of the multiple mediation models and recommended the use of the MC method. Currently, the Bayesian method has never been applied to the multiple mediation analyses. Further research is still needed on the performance of various methods, including the Bayesian method, the MC method, and the bootstrap method, in multiple mediation analyses.

## FUNDING

This research was funded by grants from the National Social Science Foundation of China [17BTJ035] and the National Natural Science Foundation of China [31771245].

## REFERENCES

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random mediating effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods*, *11*, 142–163.

Biesanz, J. C., Falk, C. F., & Savalei, V. (2010). Assessing mediational models: Testing and interval estimation for mediating effects. *Multivariate Behavioral Research*, *45*, 661–701. doi:10.1080/00273171.2010.498292

Bolin, J. H., Finch, W. H., & Stenger, R. (2018). Estimation of random coefficient multilevel models in the context of small numbers of level 2 clusters. *Educational and Psychological Measurement*. Advance online publication. doi: 10.1177/0013164418773494.

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*(2), 144-152.

Chen, J. S., 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*, *21*, 253–262. doi:10.1080/10705511.2014.882688

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Depaoli, S., & Clifton, J. P. (2015). A Bayesian approach to multilevel structural equation modeling with continuous and dichotomous outcomes. *Structural Equation Modeling*, *22*, 327–351. doi:10.1080/10705511.2014.937849

Depaoli, S., & van de Schoot, R. (2017). Improving transparency and replication in Bayesian statistics: The WAMBS-checklist. *Psychological Methods*, *22*, 240–261. doi:10.1037/met0000065

Falk, C. F., & Biesanz, J. C. (2015). Inference and interval estimation methods for mediating effects with latent variable models. *Structural Equation Modeling*, *23*, 24–38. doi:10.1080/10705511.2014.935266

Fang, J., & Zhang, M. Q. (2012). Assessing point and interval estimation for the mediating effect: Distribution of the product, nonparametric Bootstrap and Markov chain Monte Carlo methods. *Acta Psychologica Sinica*, *44*, 1408–1420. doi:10.3724/SP.J.1041.2012.01408

Fang, J., Zhang, M. Q., & Chiou, H. J. (2010). Multilevel mediation based on hierarchical linear model. *Advances in Psychological Science*, *18*, 1329–1338.

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

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed. ed.). Boca Raton, FL: CRC Press.

Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the mediating effect in statistical mediation analysis: Does method really matter? *Psychological Science*, *24*, 1918–1927. doi:10.1177/0956797613480187

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research*, *36*(2), 249–277. doi:10.1207/S15327906MBR3602_06

MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007). Distribution of the product confidence limits for the mediating effect: Program PRODCLIN. *Behavior Research Methods*, *39*, 384–389. doi:10.3758/BF03193007

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

McNeish, D. (2017). Multilevel mediation with small samples: A cautionary note on the multilevel structural equation modeling framework. *Structural Equation Modeling*, *24*, 609–625. doi:10.1080/10705511.2017.1280797

Meeker, W. Q., Cornwell, L. W., & Aroian, L. A. (1981). The product of two normally distributed random variables. In W. J. Kennedy & R. E. Odeh (Eds.), *Selected tables in mathematical statistics* (Vol. VII, pp. 129–144). Providence, RI: American Mathematical Society.

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

Muthén, L. K., & Muthén, B. O. (1998-2014). *Mplus [Computer program]*. Los Angeles, CA: Muthén & Muthén.

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*, 237–267. doi:10.1080/00273170802034844

Pituch, K. A., Stapleton, L. M., & Kang, J. Y. (2006). A comparison of single sample and bootstrap methods to assess mediation in cluster randomized trials. *Multivariate Behavioral Research*, *41*, 367–400. doi:10.1207/s15327906mbr4103_5

Pituch, K. A., Whittaker, T. A., & Stapleton, L. M. (2005). A comparison of methods to test for mediation in multisite experiments. *Multivariate Behavioral Research*, *40*, 1–23. doi:10.1207/s15327906mbr4001_1

Preacher, K. J., & Selig, J. P. (2012). Advantages of Monte Carlo confidence intervals for mediating effects. *Communication Methods and Measures*, *6*, 77–98. doi:10.1080/19312458.2012.679848

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling*, *18*, 161–182. doi:10.1080/10705511.2011.557329

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*, 209–233. doi:10.1037/a0020141

Ren, H., Wen, Z. L., Chen, Q. S., & Ye, B. J. (2013). The impact mechanism of work team leaders' psychological capital on followers' organizational citizenship behavior: A multilevel model. *Acta Psychologica Sinica*, *15*, 82–93. doi:10.3724/SP.J.1041.2013.00082

Tofighi, D., & MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods*, *43*, 692−700. doi:10.3758/s13428-010-0041-0

Tofighi, D., & MacKinnon, D. P. (2016). Monte Carlo confidence intervals for complex functions of mediating effects. *Structural Equation Modeling*, *23*, 194–205. doi:10.1080/10705511.2015.1057284

van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods*, *22*(2), 217–239. doi:10.1037/met0000100

Wang, L., & Preacher, K. J. (2015). Moderated mediation analysis using Bayesian methods. *Structural Equation Modeling*, *22*, 249–263. doi:10.1080/10705511.2014.935256

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

Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models. *Organizational Research Methods*, *12*, 695–719. doi:10.1177/1094428108327450

# APPENDIX
# MPLUS CODES IN MULTILEVEL MEDIATION ANALYSIS

```
TITLE: 2-1-1 multilevel mediation model with a random slope using
Bayesian (x = Lpsycap, m = fpsycap, y = focb)
DATA: FILE IS 3.txt;
VARIABLE:    NAMES ARE group fpsycap focb Lpsycap mfpsycap;
             USEVARIABLE = group fpsycap focb Lpsycap
mfpsycap;
             BETWEEN IS Lpsycap mfpsycap;
             within is fpsycap;
             CLUSTER IS group;
define: CENTER fpsycap (groupmean);
ANALYSIS: TYPE IS TWOLEVEL RANDOM;
             estimator = bayes;
             FBITERATIONS = 50000;
MODEL:
%WITHIN%
fpsycap focb;
b1 | focb ON fpsycap;
[fpsycap@0];
%BETWEEN%
mfpsycap focb;
mfpsycap ON Lpsycap(a);
focb ON mfpsycap(b);
focb ON Lpsycap;
MODEL PRIORS:
         a ~ n (0.41, 0.004);
         b ~ n (0.71, 0.015);
MODEL CONSTRAINT:
NEW(indb);
indb = a*b;
OUTPUT: TECH1 TECH8;
plot: type = plot2;
```