Published in final edited form as:

Struct Equ Modeling. 2017 ; 24(5): 666–683. doi:10.1080/10705511.2017.1312407.

Power in Bayesian Mediation Analysis for Small Sample Research

Milica Miočević¹, David P. MacKinnon¹, and Roy Levy²

¹Department of Psychology, Arizona State University

²T. Denny Sanford School of Social & Family Dynamics, Arizona State University

Abstract

It was suggested that Bayesian methods have potential for increasing power in mediation analysis (Koopman, Howe, Hollenbeck, & Sin, 2015; Yuan & MacKinnon, 2009). This paper compares the power of Bayesian credibility intervals for the mediated effect to the power of normal theory, distribution of the product, percentile, and bias-corrected bootstrap confidence intervals at $N = 200$. Bayesian methods with diffuse priors have power comparable to the distribution of the product and bootstrap methods, and Bayesian methods with informative priors had the most power. Varying degrees of precision of prior distributions were also examined. Increased precision led to greater power only when $N = 100$ and the effects were small, $N < 60$ and the effects were large, and $N < 200$ and the effects were medium. An empirical example from psychology illustrated a Bayesian analysis of the single mediator model from prior selection to interpreting results.

Keywords

Bayesian statistics; single mediator model; power; small sample sizes

Introduction

Mediation analysis is used to examine intermediate variables in many research areas, including psychology, sociology, communications, agriculture, political science, and epidemiology. The importance of mediating variables for basic and applied psychology has driven extensive statistical research about the best methods for hypothesis testing and interval estimation for mediated effects. Statistical simulations and analytical work indicate that methods that directly test the mediated effect and accommodate the non-normal distribution of the mediated effect are most accurate (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; MacKinnon, Lockwood, & Williams, 2004). Recently, Bayesian methods have been proposed as important complementary approaches for testing for mediation and computing the value of the mediated effect (Enders, Fairchild, & MacKinnon, 2013; Nuijten, Wetzel, Matzke, Dolan, & Wagenmakers, 2012; Wang & Preacher, 2015;

The contact author for this manuscript is Milica Miočević, Department of Methods and Statistics, Utrecht University. m.miocevic@uu.nl.

Milica Miočević, Department of Psychology, Arizona State University; David P. MacKinnon, Ph.D., Professor, Department of Psychology, Arizona State University; Roy Levy, Ph.D., Associate Professor, T. Denny Sanford School of Social & Family Dynamics, Arizona State University.

Yuan & MacKinnon, 2009) for several reasons including accuracy in small samples, probabilistic interpretation of results, and the option of using relevant prior information to improve the computation of mediated effects. The procedures and potential benefits of using prior information in Bayesian mediation analysis are the primary focus of this article. This article compares the statistical properties of Bayesian credibility intervals for the mediated effect with informative priors to the statistical properties of Bayesian credibility intervals with diffuse priors and commonly used frequentist methods for computing intervals for the mediated effect. Situations in which prior information improves statistical properties of intervals for the mediated effect are highlighted, and a detailed demonstration of specifying diffuse and informative prior distributions is provided in an empirical example.

This paper first describes the single mediator model and frequentist methods for computing interval estimates of the mediated effect. The reader is then introduced to Bayesian methods and their application to mediation analysis. Simulation studies comparing Bayesian and frequentist methods for computing intervals for the mediated effect are described next, and followed up with an empirical example where all of the methods tested in the simulation study are applied to the same data set. The paper concludes with the implications of the findings from the two simulation studies and future directions for Bayesian mediation analysis.

Single Mediator Model

The single mediator model consists of three variables: an independent variable (X) related to the mediator (M), which is related to the dependent variable (Y) (MacKinnon, 2008). This model is described using the three regression equations:

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

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

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

where c represents the total effect of X on Y, c' is the effect of X on Y adjusted for the effect of the mediator M, b measures the relation between the mediator M and the dependent variable Y adjusted for the independent variable X, and a measures the relation between X and M. i_1 , i_2 , and i_3 represent intercepts, and it is assumed that the three error terms, e_1 , e_2 , and e_3 follow a normal distribution with a mean of zero and variances σ_1^2 , σ_2^2 , and σ_3^2 , respectively (Figure 1).

Researchers can compute both point and interval estimates of the mediated effect. The most common method for calculating the point estimate of the mediated effect is as a product of

coefficients a and b in Equations 2 and 3, respectively. In addition to a point estimate, it is beneficial to also compute an interval estimate of the mediated effect (Krantz, 1999; Wilkinson, 1999). Interval estimates for the mediated effect can be computed using distribution-based methods (e.g., normal theory and distribution of the product confidence limits) or distribution-free methods (e.g., the percentile and bias-corrected bootstrap confidence limits). The product of two normal distributions, i.e., ab , is not normal (Craig, 1936; Lomnicki, 1967; Springer & Thompson, 1966), thus methods that use critical values from the normal distribution for confidence limit computation have worse statistical properties than methods that use critical values from the distribution of the product (MacKinnon, Fritz, Williams, & Lockwood, 2007; MacKinnon, Lockwood, Hoffman, West & Sheets, 2002; Tofighi & MacKinnon, 2011), and methods that make no distributional assumptions, like the bootstrap (Bollen & Stine, 1990; Efron & Tibshirani, 1993; Lockwood & MacKinnon, 1998; MacKinnon, Lockwood & Williams, 2004; Preacher & Hayes, 2004; Shrout & Bolger, 2002; Valente, Gonzalez, Mio evi, & MacKinnon, 2015). Simulation studies evaluating various interval estimates of the mediated effect have shown that overall the bias-corrected bootstrap method has the most empirical power and excessive Type I error rates in some parameter combinations (Fritz, Taylor, & MacKinnon, 2012). The percentile bootstrap and the distribution of the product have slightly less power and more accurate Type I error rates than the bias-corrected bootstrap, and normal theory confidence limit have the least power (MacKinnon, Lockwood, & Williams, 2004).

Bayesian Methods and Application to Mediation

In the frequentist framework, parameters are treated as fixed, and often the goal of a statistical analysis is to estimate their numerical value. In the Bayesian framework, the parameters are treated as random, and are assigned distributions. Prior distributions are distributions for parameters before the data are observed, and are based on available information (e.g., from a pilot study, a meta-analysis, results from a published article, or on elicited expert opinion) and factual knowledge about the parameter (e.g., a variance cannot be negative, thus priors for variance parameters should not include values below 0). A Bayesian analysis consists of updating the prior distribution with observed data using Bayes' law. The result is a posterior distribution that is then used to make inferences about parameters. Bayes' law states that $p(\theta|data)$, the posterior distribution of a collection of parameters θ , equals the product of the prior distribution for θ , $p(\theta)$, and the likelihood function, $p(data|\theta)$ divided by the marginal distribution of the data, $p(data)$, producing the equality: $p(\theta | data) = \frac{p(\theta)p(data | \theta)}{p(data)}$. The term $p(data)$ is a normalizing constant that can be omitted to obtain the following proportional relationship: $p(\theta|data) \propto p(\theta) p(data|\theta)$. This formula is at the core of how prior information is updated with observed data to produce a posterior distribution.

The available prior information can be used in "raw" (i.e., unchanged) form, or it could be calibrated (usually downweighted) in order to decrease its influence on the posterior distribution. When there is no way of distinguishing between two samples (i.e., they are drawn from the same population and data are collected using identical experimental procedures), they are said to be exchangeable. The less exchangeable the sample yielding

prior information and the focal sample, the more important it is to calibrate the prior information in order to avoid introducing bias into the statistical analysis. Some procedural guidelines for using prior information that is not exchangeable with the focal sample have been developed (de Leeuw & Klugkist, 2012; Kuiper, Buskens, Raub, & Hoijsink, 2013). Regardless of whether the researcher calibrated prior information, it is important to evaluate the impact of the prior on the inferences drawn from the posterior distribution. The evaluation of the influence of the choice of prior on the posterior distribution is called sensitivity analysis. Sensitivity to prior specifications is more pronounced in smaller samples (Gill, 2008; McNeish, 2016b), which makes sensitivity analysis more crucial in small-sample research, but it remains a useful practice even with large samples.

Sensitivity analysis is always done after the model has been estimated and the results interpreted. If the inferences from the posterior do not change notably with the various prior specifications evaluated in the sensitivity analysis, then it can be concluded that either the data carry sufficient weight in the analysis and are not overpowered by our chosen prior, and/or that the prior assumptions are sufficiently “benign” and are not adding substantial influence to the analysis. However, if the inferences from the posterior change substantially with alterations to the prior, this signals that the chosen prior carries a considerable amount of weight in the analysis relative to the observed data. If there is a strong theoretical justification for specifying a highly informative prior, then the fact that altering the prior notably changes the posterior is not sufficient reason to change this prior specification. Conversely, if a prior was chosen out of convenience and has a substantial impact on the inferences being made from the posterior, then the results of the sensitivity analysis suggest this prior ought to be reconsidered (Gill, 2008).

In the case of the mediated effect, it is possible to assign priors to either parameters in the regression models for M and Y (Yuan & MacKinnon, 2009), or to the covariance matrix of X, M, and Y (Enders, Fairchild, & MacKinnon, 2014); these will be referred to as the method of coefficients and the method of covariances, respectively (Mio evi & MacKinnon, 2014). Prior distributions are often labeled as either diffuse, weakly-informative, or informative, however, these are false categories as the level of informativeness of a prior distribution exists on a continuum. The parameters of a prior distribution are called hyperparameters. In the method of coefficients the intercepts i_2 and i_3 , and the regression coefficients a , b , and c' in Equations 2 and 3 are assigned normal prior distributions, $N(\mu, \sigma^2)$, with the mean hyperparameter equal to the expected value of the coefficient in the population, and the variance (or precision, $1/\sigma^2$) hyperparameter reflecting the researcher's level of confidence that the expected value is accurate. In normal priors the level of informativeness is determined by the choice of the variance (or precision) parameter. A higher variance (i.e., lower precision) hyperparameter reflects less confidence that the prior expectation is equal to the value of the parameter, and thus creates a more diffuse prior for the regression coefficient. Decreasing the variance hyperparameter in the normal prior (i.e., increasing the precision hyperparameter) reflects more confidence in the specified mean hyperparameter, and thus leads to a more informative prior. Another possibility for prior distributions for regression coefficients and intercepts is the uniform distribution, $U(a, b)$, with hyperparameters equal to the expected range of values for the coefficient (Yuan & MacKinnon, 2009). Uniform priors are a good choice when the possible range of a

regression coefficient or intercept is known, e.g., the intercept cannot be outside of the range of the scale of the dependent variable (Levy & Crawford, 2009).

Residual variances in regression analysis are usually assigned inverse-gamma prior distributions. The inverse-gamma (IG) is said to be a conditionally conjugate prior for the variance parameter, meaning that if an inverse gamma prior is assigned to the variance parameter, then the posterior for the variance, conditional on other model parameters, is also an inverse-gamma (Gelman, 2006). The inverse-gamma distribution is specified using the shape (α), and scale (β) hyperparameters. Because the hyperparameters of an inverse-gamma prior are not as easy to conceptualize as the hyperparameters of a normal prior, one can specify α and β by imagining a prior sample was observed. Using this approach, α equals one half of the variance of the prior pseudo-sample, and β equals one half of the product of the variance of the prior pseudo-sample and the sample size of the pseudo-sample (Gelman, Carlin, Stern, & Rubin, 2004). The simulation studies in this project made use of normal priors for regression coefficients and intercepts, and inverse-gamma priors for residual variances.

Another approach for Bayesian mediation is called the method of covariances, and consists of specifying an inverse-Wishart prior, $p(\Sigma) \sim W^{-1}(\nu, S)$ for the covariance matrix of X, M, and Y (Enders, Fairchild, & MacKinnon, 2014). The hyperparameters of an inverse-Wishart are the degrees of freedom parameter, ν , and a symmetric, positive definite matrix S. The degrees of freedom parameter ν is akin to the prior sample size, and the S matrix is often set to the matrix of sums of squares and cross-products. The inverse-Wishart with a low value of ν (at least equal to the number of variables so the distribution does not have a degenerate form) and a fixed scale matrix S is used as a non-informative proper prior for the covariance matrix (Daniels & Kass, 1999). Increasing ν makes the inverse-Wishart prior more informative. After the specification of a prior and the likelihood function, Markov chain Monte Carlo (MCMC) methods are used to approximate the posterior distribution. The mediated effect is computed as the product of coefficients a and b at each MCMC iteration. In the method of covariances the mediated effect is computed from the elements of the

posterior covariance matrix as $ab = \frac{s_{XM}(s_{MY}s_X^2 - s_{XM}s_{XY})}{s_X^2(s_X^2s_M^2 - s_{XM}^2)}$ (Enders, Fairchild, & MacKinnon,

2014; MacKinnon, 2008). The posterior for ab is summarized with a point summary (e.g., mean or median), and/or with an equal-tail and/or a highest posterior density (HPD) credibility interval. Equal-tail credibility intervals are constructed by selecting the $(\alpha/2)^{\text{th}}$ and $(1-\alpha/2)^{\text{th}}$ percentiles of the posterior ($1-\alpha$ is the nominal coverage). HPD intervals are constructed so no values outside the interval have higher probability than values within the interval (Gelman et al., 2004).

There are several benefits to doing mediation analysis in the Bayesian framework: the summaries of the posterior have probabilistic interpretations, thus one can compute the 95% probabilities that ab lies between two values and/or that $ab = 0$; prior information can be explicitly and transparently incorporated into the analysis, which lowers the standard deviation of the posterior, and decreases interval width of credible intervals (van de Schoot, et al., 2014). Furthermore, in the frequentist framework, power and Type I error rate vary

together: as one increases, so does the other. In other words, setting nominal α to 0.10 instead of 0.05 increases power because with larger α the criterion for rejecting the null hypothesis is less strict (i.e., effects with larger p -values will be considered significant). This is desirable when there is an effect in the population, but problematic because larger nominal α also means higher Type I error rates when there is no effect in the population. The bias-corrected bootstrap is the frequentist method with the most power to detect the mediated effect, but it also has the highest Type I error rates, which is why it is not always recommended (Fritz, Taylor, & MacKinnon, 2012; MacKinnon, Lockwood, & Williams, 2004). With Bayesian methods, meaningful information can still be obtained from a study without having to sacrifice power in order to control Type I error rates (van de Schoot, Hoijtink, Mulder, Van Aken, Orobio de Castro, Meeus et al., 2011). Even though Bayesian methods have been suggested as a method for increasing power in mediation analysis with small samples (Yuan & MacKinnon, 2009), there is presently no study comparing power using Bayesian methods to power of frequentist methods such as normal theory, distribution of the product, percentile, and bias-corrected bootstrap.

Goals of the Study

There is a great need for methods suited for small sample research (Etz & Arroyo, 2015). Several frequentist methods for mediation analysis have been designed to deal with issues that arise in small samples (Hoyle & Kenny, 1999), and Bayesian methods have been proposed as an alternative to frequentist methods with small samples for SEM (Hox, van de Schoot, & Matthijsse, 2012; Lee & Song, 2004), growth mixture models (Depaoli, 2013), and multilevel models (McNeish, 2016a). It was suggested that Bayesian methods have potential for increasing power in mediation analysis (Koopman, Howe, Hollenbeck, & Sin, 2015; Yuan & MacKinnon, 2009), however, this is the first paper to focus on the potential of Bayesian methods with diffuse and informative conjugate priors for mediation analysis in small samples. The first goal of this study is to examine the extent to which using prior information in a Bayesian analysis with small samples increases the power to detect the mediated effect compared to commonly used frequentist methods and Bayesian methods with diffuse priors. It is intuitive that using accurate priors will lead to higher power, less Type I error rate, higher coverage, lower imbalance, and smaller interval width compared to using diffuse priors or no prior information. However, the amount of improvement in statistical properties of intervals for ab with accurate priors is unknown. The first simulation study addresses this question in addition to comparing Bayesian methods with diffuse priors to the best available frequentist methods. The second simulation study evaluates the changes in power to detect the mediated effect with gradual increases in the precision of the prior. Another goal of this paper is to give a detailed example of how to select priors in a Bayesian analysis with existing prior information for the method of coefficients and the method of covariances, how to interpret the results, and conduct a sensitivity analysis. The paper concludes with recommendations for substantive applications of Bayesian mediation analysis.

Simulation Studies

Study 1

The goal of the first simulation study was to examine the amount of increase in empirical power of Bayesian equal-tail credibility intervals with accurate informative priors relative to empirical power of Bayesian equal-tail credibility intervals with diffuse priors and commonly used frequentist methods for interval estimation, i.e., normal theory, distribution of the product, percentile bootstrap, and bias-corrected bootstrap confidence limits in samples up to 200. Additionally, all interval estimators were compared on Type I error rate evaluated using Bradley's (1978) robustness criterion (i.e., values between 0.025 and 0.075 are considered adequate), coverage evaluated using Bradley's (1978) robustness criterion (i.e. values between 0.925 and 0.975 are deemed adequate), interval width (lower widths indicate more informative intervals), and imbalance calculated as the difference between the number of true values that fall to the right versus the left side of the interval (the closer to zero, the better).

Methods—Population values for parameters were chosen based on conditions where there was the most discrepancy between methods in confidence limit estimation in prior research (MacKinnon, Lockwood & Williams, 2004). This yielded thirteen parameter combinations: $a=0$ $b=0$ $c'=0$, $a=0$ $b=0.14$ $c'=0$, $a=0$ $b=0.39$ $c'=0$, $a=0$ $b=0.59$ $c'=0$, $a=0.14$ $b=0.14$ $c'=0$, $a=0.39$ $b=0.39$ $c'=0$, $a=0.59$ $b=0.59$ $c'=0$, $a=0.14$ $b=0.39$ $c'=0$, $a=0.14$ $b=0.59$ $c'=0$, $a=0.39$ $b=0.59$ $c'=0$, $a=0.14$ $b=0.14$ $c'=0.39$, $a=0.39$ $b=0.39$ $c'=0.39$, $a=0.59$ $b=0.59$ $c'=0.39$. Two values of the c' coefficient (0 and 0.39) were included in the simulation in order to determine whether performance of interval estimates differs for complete ($c'=0$) and incomplete ($c' \neq 0$) mediation models.

Random samples of sizes $N = 20, 40, 60, 100$, and 200 were obtained from each population and 95% confidence limits for the mediated effect were calculated using normal theory, the distribution of the product, percentile bootstrap, and bias-corrected bootstrap. Four Bayesian methods were evaluated: the method of coefficients with diffuse prior distributions, the method of coefficients with informative prior distributions, the method of covariances with diffuse prior distributions, and the method of covariances with informative prior distributions. The informative prior distributions for the methods of coefficients and covariances were calculated using population-generating values from the simulation study; for simplicity, this specification of priors will be referred to as the “truth” about the parameters.

Method of coefficients with diffuse prior distributions: Diffuse normal prior distributions centered at the true values of coefficients a , b , and c' with variance hyperparameters equal to 10^3 were assigned to the regression coefficients, and residual variances σ^2_M and σ^2_Y were assigned inverse-gamma distributions with both hyperparameters equal to .01 (so that the expectation of this distribution is 1, which was the true residual variance in the simulation).

Method of coefficients with informative prior distributions: Normal prior distributions centered at the true values of coefficients a , b , and c' with standard deviations equal to the standard errors of respective coefficients calculated using simulated (population-generating)

values at a given sample size were assigned to the regression coefficients, and residual variances σ^2_M and σ^2_Y were assigned inverse-gamma distributions with both hyperparameters equal to .01, thus yielding an accurate specification of the prior for the residual variance.

Method of covariances with a diffuse prior distribution: The vector of means of X, M, and Y was assigned a multivariate normal distribution, and the covariance matrix was assigned an inverse-Wishart distribution with 3 degrees of freedom. The S matrix (matrix of sums of squares of the variables on the diagonal and covariations between pairs of variables as the off-diagonal elements) of the inverse-Wishart prior was specified so that the prior expectations for variances and covariance between variables were 1 and 0, respectively.

Method of covariances with an informative prior: The vector of means of X, M, and Y was assigned a multivariate normal prior distribution with means of 0, variances of 1000, and covariances of 0. The covariance matrix was assigned an inverse-Wishart prior distribution with the degrees of freedom hyperparameter equal to N_p , the size of the observed sample in the condition (e.g., if the sample size in the condition was 60, the degrees of freedom parameter of the inverse-Wishart was also set to 60). The S matrix was specified so that the prior expectations for each variance and covariance equal their respective true values.

For Bayesian methods the number of burn-in iterations, the thinning parameter, and the number of retained iterations were selected based on preliminary analyses of convergence using trace plots and plots of autocorrelation. For the method of coefficients the number of burn-in iterations was equal to 1000 (the default in SAS PROC MCMC), the thinning parameter was set to 5, and the number of MCMC draws after burn-in was set to 5000 thus leaving 1000 iterations to approximate the posterior after thinning (SAS Institute Inc., 2013). The same specifications were used for the method of covariances. In order to test whether 1000 iterations are sufficient for approximating the posterior for this model, three parameter combinations were rerun with 10000 retained iterations and the results were unchanged. The simulation consisted of 1000 replications for each combinations of effect size and sample size. Average empirical power, coverage, Type I error rate, imbalance, and interval width were obtained for all eight methods.

Results—Previous findings from the literature on frequentist methods were confirmed, i.e., the bias-corrected bootstrap had the highest power, followed by the percentile bootstrap and the distribution of the product, and normal theory confidence limits had the lowest power. As expected, Bayesian methods with informative prior distributions had the highest power of all methods at all sample sizes, and Bayesian methods with diffuse priors had average power comparable to the distribution of the product and percentile bootstrap confidence limits. The beneficial effects of accurate informative priors on empirical power were inferred from the examination of the shape of the power curves as a function of sample size for all methods at different parameter combinations (Figure 2). With medium and large a and b , the values of power for Bayesian methods with informative prior distributions plateaued at $N = 40$, meaning that past this point there were no substantial increases in power with increases in sample size. The power curves of the remaining six methods started plateauing at $N = 100$ if at least

one of the coefficients (a or b) is 0.59. Thus, when effects are medium or large, $N = 40$ is sufficient for Bayesian methods with informative priors to attain the highest value for power, whereas for Bayesian methods with diffuse priors and frequentist methods this happens at $N = 200$ when at least one of the mediation paths (a or b) is medium and at $N = 60$ when both $a = b = 0.59$. The value of c' did not change the power profile of any of the methods.

There were four combinations where the mediated effect was zero. All eight methods had Type I error rates lower than the nominal 0.05 level at all sample sizes. The bias-corrected bootstrap had the highest Type I error rates, whereas the Bayesian methods with informative priors had almost no Type I error regardless of sample size or the size of the b coefficient. Bayesian methods with diffuse prior distributions had Type I error rates below 0.05 for all combinations of b and sample size except when b was medium or large and $N = 200$.

Bayesian methods with diffuse and informative priors and the distribution of the product method had coverage above the nominal level at all sample sizes. Normal theory confidence limits and the two bootstrap methods had coverage levels closest to the nominal level of 0.95. Bayesian methods with informative prior distributions had coverage close to 1 regardless of sample size or effect size. The bias-corrected bootstrap had the lowest coverage for $a=0$ and $b>0$, which is consistent with the findings that the bias-corrected bootstrap had higher Type I error than other methods for $a=0$ and $b>0$. Bayesian methods with informative prior distributions had the lowest values of interval width and imbalance closest to 0 regardless of sample size and effect size. The size of the c' coefficient did not have an impact on the coverage, power, imbalance, and interval width of the intervals for the mediated effect for any of the methods, as determined by non-significant t-test that compared statistical properties for combinations with $a=b$ with $c'=0$ and $a=b$ with $c'=0.39$.

In summary, the results of Study 1 suggest that Bayesian methods with accurate prior specifications are an excellent way to increase power when the sample size is below 200, without the risk of excess Type I error rates, lower coverage, higher interval width, and higher imbalance. Even with diffuse priors Bayesian methods still performed comparably to the best frequentist methods, i.e., the distribution of the product and the percentile bootstrap. However, priors with hyperparameters computed using population values of parameters and trustworthy enough to be given the same weight in the analysis as the focal sample are the best case scenario, and may be rare in practice. Keeping in mind that substantive researchers may not want to give the prior and likelihood equal weight, Study 2 was designed to examine the amount of precision in the prior for the method of coefficients that is necessary in order to achieve the desirable interval properties of Bayesian methods with informative priors observed in Study 1.

Study 2

Study 2 examined if the observed increases in power with informative prior distributions using the method of coefficients are still present with precision parameters in the prior distributions for regression coefficients that are different than precisions calculated from true parameters (as was done in Study 1). Study 2 also investigated whether more prior information about the a -path or the b -path led to greater increases in power for the mediated effect.

Methods—Study 2 examined the effect of precision in the normal priors for regression coefficients a and b on changes in power to detect the mediated effect for the Bayesian method of coefficients. Populations with the following values for parameters a , b , and c' were simulated: $a=b=0.14$ with $c'=0$, $a=b=0.39$ with $c'=0$, $a=b=0.59$ with $c'=0$, $a=b=0.14$ with $c'=0.39$, $a=b=0.39$ with $c'=0.39$, and $a=b=0.59$ with $c'=0.39$. Random samples of sizes $N = 20, 40, 60, 100$, and 200 were drawn from each population and 95% credibility intervals for the mediated effect were calculated using the method of coefficients with different priors for regression coefficients a and b . Ten conditions were evaluated in the study: five where the precision parameter in the prior for the coefficient b was set to 10^{-3} and the precision parameter in the prior for coefficient a was varied ($10^{-2}, 10^{-1}, 10^0, 10^1, 10^2$), and five where the precision parameter in the prior for the coefficient a was set to 10^{-3} and the precision parameter in the prior for coefficient b was varied ($10^{-2}, 10^{-1}, 10^0, 10^1, 10^2$). The true precisions for a and b are equal to 1 divided by the true variance (computed using formulas on pp. 88–89 of MacKinnon, 2008) of the respective parameter at each sample size. The true variance (and precision) of the coefficient depends on effect sizes and sample size. Thus, a fixed value of the precision hyperparameter in the prior for a or b ($10^{-2}, 10^{-1}, 10^0, 10^1, 10^2$) is a different fraction (percentage) of the true precision of the coefficient at different values of effect size and sample size.

Table 1 shows how the values of precision relate to the true precisions of parameters a and b at each sample size and for each precision parameter. The entries in Table 1 are the percentages of the true precision values for parameter values examined in Study 2; percentages below 100% indicate that the prior was more diffuse than the corresponding prior in Study 1 (recall that the prior in Study 1 had a precision parameter equal to the inverse of the true variance of the corresponding regression coefficient), and percentages above 100% indicate a more informative prior than the corresponding prior in Study 1. For example, when the precision hyperparameter of the prior for regression coefficient a is set to 10 and the sample size is 100, the precision hyperparameter of 10 is 10.2% of the true precision of a . The values of precision examined in Study 2 ($10^{-2}, 10^{-1}, 10^0, 10^1, 10^2$) are generally lower than the true precisions of each parameter, except with the precision parameter equal to 100 and N between 20 and 100.

In Study 2 the mean hyperparameters in the prior distributions for the regression coefficients a and b were always set equal to the true value. In all ten conditions, c' was assigned a normal prior distribution with the true value as the mean hyperparameter and a precision hyperparameter of 10^{-3} . The intercepts in the regression equations predicting M and Y were assigned normal prior distributions with mean hyperparameters of 0 and precision hyperparameters equal to 10^{-3} . The priors for the residual precision of M and Y were the same for all conditions (gamma distributions with shape and inverse scale parameters equal to .01), as in Study 1. The simulation consisted of 1000 iterations per combination of effect size and sample size. The only change from the simulation in Study 1 was that the thinning parameter was set to 1, so the posterior distributions were approximated using 5000 draws instead of 1000. Average empirical power for all six conditions was computed for $N = 20, 40, 60, 100$, and 200 .

Results—Study 2 examined the extent to which the observed increases in power are still present with precision hyperparameters in normal priors for regression coefficients equal to powers of 10, and whether more prior information (more precision) for the a -path or the b -path led to greater increases in power for the mediated effect. The relation between precision and power can be conceived in two ways. First, increases in power can be observed as a function of the values of the precision parameter for a and b , which were 10^{-2} , 10^{-1} , 10^0 , 10^1 , and 10^2 . Increases in power can also be conceived of as a function of increases in precision equated across sample size; that is, increases in power as a function of the percentage of the true precision that was encoded in the prior distribution. Findings indicated that increasing precision tenfold affects power differently depending on the sample size and effect size. Increases in the precision hyperparameter produced the steepest increase in power for medium effects, some increase in power for large effects, and no increase in power for small effects. More specifically, increases in precision led to greater power only when the sample size was at least 100 and the effects were small, when sample size was below 60 and the effects were large, and the most notable increases in power were when $N < 200$ and the effects were medium. When looking at the relation of the precision in the prior to the true precision of the parameter (which is unknown to the researcher), and the extent to which prior precisions larger than the true precision of the coefficient led to more power, the conclusion was similar: increases in precision relative to the true precision may not always lead to noticeable increases in power. Furthermore, there was no differential increase in power depending on whether there was more precision in the prior for a versus the prior for b . Finally, t -tests used to compare power in conditions with $c' = .39$ and conditions with $c' = 0$ demonstrated no difference in power to detect the mediated effect at different values of c' .

The most notable finding in Study 2 was that for certain combinations of parameter values and sample size, there are ceiling effects (power cannot be larger than 1) and floor effects (power cannot be smaller than 0), and increasing or decreasing the precision parameter of the prior distribution by a factor of 10 for either the a or b coefficient can have almost no impact on power (Figure 3). For small effects and small sample sizes, and for large effects and $N = 100$ there was almost no change in power as a consequence of increasing the precision parameter.

Empirical Example—Data for the empirical example come from two exchangeable studies of word recall in a sample of students from a university in the United States (MacKinnon, Wurpts, & Valente, 2014). The second study provided the focal data set for the analysis, and the results of the analysis of the first study were only used as a source of prior information. This scenario emulates the real life situation in which substantive researchers have only published results to use as prior information for their study. It was hypothesized that individuals who make images for words remember more words. To test this hypothesis, students were divided into two groups (the imagery group and the repetition group) that received different instructions. Students assigned to the repetition condition were instructed to repeat the words they hear, while students in the imagery condition were instructed to make images for the words. Students then heard a list of 20 words, and subsequently reported the extent to which they made images for words (on a scale from 1-not at all to 9-

absolutely), and wrote down as many words from the list as they remembered. The independent variable was condition (X), the mediator was the extent to which students made images for words (M), and the dependent variable was the number of words recalled (Y). Intervals for the mediated effect of condition (X) on number of words recalled (Y) through imagery (M) were computed using normal theory, distribution of the product, percentile bootstrap, bias-corrected bootstrap, Bayesian methods with diffuse priors, and Bayesian methods with informative priors (Table 2).

The diffuse priors for the Bayesian analysis were selected based on the scales of the variables and examples in methodological literature of what is defined as diffuse. Given the type of dependent variables (M being imagery on a 1–9 scale, and Y being the number of words recalled out of 20) the regression coefficient a was expected to be smaller than 8 in absolute value and regression coefficients b and c' were expected to be smaller than 5 in absolute value, thus for the method of coefficients the variance hyperparameter in diffuse normal priors for the regression coefficient had to be large enough to make values between -8 and 8 and -5 and 5 (respectively) almost equally likely. Keeping this in mind, regression coefficients a , b , and c' were assigned normal priors with the mean hyperparameter equal to 0 and the variance hyperparameter equal to 10^3 . The intercepts in the regression equations predicting M and Y could not be below or above the limits of the scales of the variables, thus in the diffuse case where there are no assumptions of any one value being more likely than the other values both intercepts were assigned uniform priors with bounds equal to the limits of the scale (i.e., 1 – 9 for M, and 0 – 20 for Y). In the diffuse case for the method of covariances, the matrix of means of X, M, and Y was assigned a multivariate normal prior with a vector of 0.5, 5, and 3 as the mean hyperparameter and $\mathbf{I} \times 1000$ (a diagonal matrix with 1000 as each of the non-zero elements) as the variance hyperparameter. This specification reflects the assumption that the mean hyperparameters of the priors for X and M were equal to the mean values of their respective scales (X was a dummy variable, and M ranged from 1 to 9), and the assumption that the mean of the prior for Y was equal to 3. Given that Y was the number of words recalled ranging from 0 to 20, this choice of mean hyperparameter reflected an assumption that on average participants recalled very few words relative to the number of words they heard. The variance hyperparameter of the prior for the mean vector of X, M, and Y indicated very little confidence in the chosen mean hyperparameter values for X, M, and Y, thus making the prior distribution for the mean vector diffuse. The covariance matrix of X, M, and Y was assigned an inverse-Wishart prior with 4 degrees of freedom (number of variables in the model plus 1; akin to saying the prior data set consisted of 4 participants) and with a S matrix with zeros as the non-diagonal elements, thus representing the assumption of no covariances between variables, and with values of 0.065, 3, and 3 as the diagonal elements thus encoding the assumption of the variances for X, M, and Y being equal to 0.025, 1, and 1 (respectively).

In order to select values for the multivariate normal and inverse-Wishart priors for the method of coefficients, expected means of X, M, and Y were elicited first, and then the question of likely values around these mean values was posed. Standard deviations around the means of X, M, and Y were chosen so that highly unlikely values for X, M, and Y corresponded to the mean value ± 3 standard deviations. The expected variances of X, M,

and Y were then computed by squaring the elicited standard deviations, and the S matrix for the inverse-Wishart prior was computed by multiplying each element of the covariance matrix by the number of participants from the pseudo-prior sample (corresponding to the degrees of freedom parameter of the inverse-Wishart, here 4) minus 1, which was equal to 3. The procedure for selecting the hyperparameters of the diffuse priors was inspired by guidelines on eliciting prior information from experts (Low Choy, O'Leary, & Mengersen, 2009; Kadane & Wolfson, 1998; O'Hagan, 1998).

Since we had no reason to consider participants in the prior sample to be any different from the participants in the observed sample and the methods in the two experiments were identical, the findings from the first data set were not downweighted for the creation of prior distributions for the analysis of the second data set. Informative priors were created using raw regression coefficients and standard errors as the mean and standard deviation hyperparameters of the priors for the regression coefficients and intercepts in the method of coefficients, and the observed covariation matrix and sample size ($N_{\text{prior}} = 35$) were used as hyperparameters of the inverse-Wishart prior for the covariance matrix in the method of covariances.

The frequentist intervals for the mediated effect ab were computed using SAS for Windows Version 9.4 and RMediation (Tofighi & MacKinnon, 2011), and Bayesian credibility intervals were computed using PROC MCMC (Mio evi & MacKinnon, 2014) and the % postint macro (SAS Institute Inc., 2013). Furthermore, MCMC draws were also used to compute the probability that the mediated effect ab is greater than 1, greater than 2, and greater than 5; these are the probabilities that by being in the imagery instead of the repetition condition the participants remembered at least one more, two more, or five more words through making images for the words. SAS syntax for the Bayesian analyses is available in the Appendix.

None of the intervals constructed using frequentist and Bayesian methods contained zero, thus if confidence and credibility intervals were used for significance testing (as is common in the literature), one would conclude that the indirect effect of condition on words recalled through imagery was statistically significant (Table 3). The credibility intervals produced using Bayesian methods with informative priors were narrower and thus more precise than intervals produced using Bayesian methods with noninformative priors and frequentist methods; this observation is consistent with the results of the first simulation study. The probability that the participants in the imagery condition remembered at least one more word by making images was 91% for the method of coefficients and 92% for the method of covariances with diffuse priors, and 99% for both methods with informative priors. The probability that the participants remembered at least 2 more words by making more images for words if they were in the imagery condition was 56% for Bayesian methods with diffuse priors and 80% with informative priors. Regardless of the Bayesian method and prior specification, the probability of remembering at least 5 more words by making more images for words after being in the imagery condition was always 0.

A sensitivity analysis evaluated the effects of the specification of informative priors on the observed results. In order to evaluate the relative impact of the chosen priors on the observed

results, the informative priors were made even more informative and also less informative. For the method of coefficients, the calibration was carried out by increasing and/or decreasing the spread hyperparameter of the priors for regression coefficients. SAS PROC MCMC allows for the spread of a normal distribution to be quantified in the variance, precision, or standard deviation metric. For all of the analyses of the empirical example, including the sensitivity analysis, the standard deviation metric was used as it is the quantity that is most often numerically equivalent to the standard error (i.e., estimates have standard errors due to repeated sampling, while the standard deviation of a prior or posterior distribution is a measure of the spread of the distribution that does not invoke the assumption of repeated sampling) and all software packages for regression analysis report the estimated coefficient and its standard error, which can then be converted into a normal prior with the mean equal to the estimated regression coefficient and the standard deviation of the prior equal to the standard error of the estimated regression coefficient. In the sensitivity analysis the standard deviation parameter of the priors for the regression coefficients was varied to equal half of, twice, five times, and ten times the observed standard error of the corresponding regression coefficient. There was no reason to think that the regression coefficients would introduce bias into the analysis since the prior and the focal sample are exchangeable, thus only the spread of the prior distribution was varied.

The goal of the sensitivity analysis in this case was to evaluate whether using the uncalibrated estimates from the prior study as prior information had substantial influence on the conclusions about the mediated effect in the focal sample. In some situations it might be of interest to vary the mean hyperparameter of the normal priors for regression coefficients, e.g., when prior distributions are constructed from a prior sample that is not drawn from the same population and/or when the data were collected using different measurement instruments and procedures, or when a subject matter expert thinks there are several plausible values the regression coefficient could take. In our case, the regression coefficients from the prior sample were assumed to be unbiased, and the concern was whether the prior information should carry the weight of $N = 35$ (i.e., the actual size of the prior sample), or if it should be downweighted in order to assume a smaller (and consequently less influential) prior sample.

In the method of covariances, the informativeness of the prior distribution is set with the choice of the degrees of freedom parameter; the higher the degrees of freedom parameter, the more informative the inverse-Wishart prior distribution. In the sensitivity analysis for the method of covariances the degrees of freedom parameter was varied to equal twice, half, a fifth, and a tenth of the observed prior sample size, corresponding to the assumption of the prior information coming from 70, 17.5, 7, and 3.5 participants, respectively. The choice to multiply the degrees of freedom parameter by 2 and then divide it by 2, 5, and 10 was made in order to keep the sensitivity analyses for the method of covariances and the method of coefficients consistent. Note that this choice resulted in assuming a prior sample size of 17.5 and 3.5 participants; such a specification is not problematic in software, nor does it result in an improper prior distribution as the degrees of freedom parameter was larger than the number of variables in the model. Priors in the sensitivity analyses and the 95% credibility intervals for *ab* are summarized in Table 4.

The sensitivity analysis examined the effect of prior specification on the values of upper and lower credibility limits. The results indicated that with more diffuse priors the intervals for the mediated effect became wider, as expected, however, even with priors that were five or ten times more diffuse relative to the available prior information, both the equal-tail and HPD credibility intervals for ab did not include zero and included the values of 1, 2, and 3 thus indicating that it is possible that being in the imagery instead of the repetition condition resulted in at least one more word remembered through making images for words.

Discussion

This research examined the increase in power with the use of accurate prior information in a Bayesian mediation analysis relative to Bayesian methods with diffuse priors and commonly used frequentist methods, i.e., normal theory confidence limits, distribution of the product, and the percentile bootstrap. Findings indicate that including accurate prior information produces intervals with higher power while maintaining satisfactory levels of Type I error rate, coverage, interval width, and imbalance, but not for every parameter and sample size combination. The power gains from using Bayesian methods depend on the sample size, effect size, and the informativeness of the prior. With diffuse prior distributions, Bayesian credibility intervals for the mediated effect have comparable statistical properties to intervals obtained using the distribution of the product and the percentile bootstrap. Even in the absence of relevant prior information Bayesian methods are still useful because credibility intervals for the mediated effect have a probabilistic interpretation, and one can compute the probability that the mediated effect is as large as or larger than a substantively meaningful value (e.g., one, two, and five words).

The simulation studies examined informative priors that contain accurate information. When analyzing real data, as was done in the empirical example, it is not known whether the chosen priors are accurate. Thus, the recommendation for substantive researchers is to consider making the priors less informative than the available amount of prior information using procedures illustrated in the empirical example, i.e., by making the standard deviations of the priors larger than the standard error of the regression coefficient in the prior study in the method of coefficients, and by making the degrees of freedom parameter of the inverse-Wishart smaller than the size of the prior sample in the method of covariances.

If the same information is encoded in the informative priors for the method of coefficients and the priors for the method of covariances, the two ought to yield the same numerical summaries of the posterior distribution. In the simulation for Study 1 of this project the informative priors for the method of coefficients carried less information than the informative priors for the method of covariances. This was the case because the univariate priors for the residual variances of M and Y in the informative priors condition were still diffuse (i.e., inverse-gamma priors with $\alpha=\beta=.01$, thus yielding an accurate expectation of the residual variance that is equal to 1, the value of the parameter in the simulated population, but carrying the weight of .02 “observations”). On the other hand, the informative priors for the method of covariances were informative for every element of the covariance matrix (because the df hyperparameter of the inverse-Wishart prior communicates the informativeness of the prior for the entire matrix, and the df

hyperparameter for informative priors was set to be of the same size as the N in the focal sample yielding the likelihood function). This choice was made to emulate a situation where researchers find published results about regression coefficients but not residual variances, and about the entire covariance matrix along with the prior sample size.

One possible limitation of the simulation studies in this paper is that they examined only equal-tail credibility intervals. Given that the product of two normal variables (e.g., coefficients a and b) is not normal (Craig, 1936; Lomnicki, 1967; Springer & Thompson, 1966), the mediated effect is summarized better with intervals that allow for asymmetry in the distribution of the parameter (MacKinnon, Fritz, Williams, & Lockwood, 2007; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; MacKinnon, Lockwood, & Williams, 2004). Higher posterior density (HPD) credibility intervals would accommodate this asymmetry better, and their statistical properties would be at least as satisfactory as those of equal-tail credibility intervals.

This project examined the potential for increasing power in the single mediator model using Bayesian methods with informative prior distributions. As expected, including accurate prior information in the analysis leads to superior interval properties of credibility intervals for the mediated effect relative to credibility intervals obtained using diffuse priors and confidence intervals. However, the statistical benefits of the inclusion of prior information depend on the combination of effect sizes and sample size. The results of the simulation studies in this paper are primarily generalizable to situations with the same scales for variables X , M , and Y , same sample sizes, same effect sizes, and same prior specifications. Changes in any of those factors may lead to different conclusions. This paper also offered a detailed example of how to conduct a Bayesian mediation analysis, starting with prior selection and ending with a sensitivity analysis. Bayesian mediation analysis is a growing field, and there are many questions regarding the appropriate use of prior information in mediation models that have yet to be answered, such as questions concerning procedures for calibrating prior information from different sources and the impact inaccurate priors have on the results (Depaoli, 2014).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported in part by National Institute on Drug Abuse, Grant No. R01DA009757. The authors would like to thank Stephen G. West and Craig K. Enders for their feedback during the development of the research presented in this manuscript.

References

- Bradley JV. Robustness? British Journal of Mathematical and Statistical Psychology. 1978; 31(2):144–152.
- Craig CC. On the frequency function of xy . Annals of Mathematical Statistics. 1936; 7:1–15.
- Daniels MJ, Kass RE. Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. Journal of the American Statistical Association. 1999; 94(448):1254–1263.

- Depaoli S. Mixture class recovery in GMM under varying degrees of class separation: Frequentist versus Bayesian estimation. *Psychological Methods*. 2013; 18(2):186–219. [PubMed: 23527607]
- Depaoli. The Impact of Inaccurate “Informative” Priors for Growth Parameters in Bayesian Growth Mixture Modeling. *Structural Equation Modeling*. 2014; 21(2):239–252.
- Enders CK, Fairchild AJ, MacKinnon DP. A Bayesian approach to estimating mediation effects with missing data. *Multivariate Behavioral Research*. 2013; 48(3):340–369. [PubMed: 24039298]
- Etz KE, Arroyo JA. Small Sample Research: Considerations Beyond Statistical Power. *Prevention Science*. 2015; 16(7):1033–1036. [PubMed: 26281902]
- Fritz MS, Taylor AB, MacKinnon DP. Explanation of two anomalous results in statistical mediation analysis. *Multivariate Behavioral Research*. 2012; 47:61–87. [PubMed: 24049213]
- Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*. 2006; 1:515–533.
- Gelman, A., Carlin, JB., Stern, HS., Rubin, DB. *Bayesian data analysis*. CRC press; 2004.
- Gill, J. *Bayesian methods: A social and behavioral sciences approach*. CRC press; 2008.
- Hox JJ, van de Schoot R, Matthijsse S. How few countries will do? Comparative survey analysis from a Bayesian perspective. *Survey Research Methods*. 2012; 6(2):87–93.
- Hoyle, RH., Kenny, DA. Sample size, reliability, and tests of statistical mediation. In: Hoyle, RH., editor. *Statistical strategies for small sample research*. Thousand Oaks, CA: Sage; 1999. p. 195–222.
- Kadane JB, Wolfson LJ. Experiences in elicitation. *The Statistician*. 1998; 47:3–19.
- Koopman J, Howe M, Hollenbeck JR, Sin HP. Small sample mediation testing: misplaced confidence in bootstrapped confidence intervals. *Journal of Applied Psychology*. 2015; 100(1):194–202. [PubMed: 24731180]
- Krantz DH. The null hypothesis testing controversy in psychology. *Journal of the American Statistical Association*. 1999; 94(448):1372–1381.
- Kuiper RM, Buskens V, Raub W, Hoijtink H. Combining statistical evidence from several studies a method using Bayesian updating and an example from research on trust problems in social and economic exchange. *Sociological Methods & Research*. 2013; 42(1):60–81.
- Lee SY, Song XY. Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research*. 2004; 39(4):653–686. [PubMed: 26745462]
- de Leeuw C, Klugkist I. Augmenting data with published results in Bayesian linear regression. *Multivariate Behavioral Research*. 2012; 47(3):369–391. [PubMed: 26814603]
- Levy R, Crawford AV. Incorporating substantive knowledge into regression via a Bayesian approach to modeling. *Multiple Linear Regression Viewpoints*. 2009; 35:4–9.
- Lomnicki ZA. On the distribution of products of random variables. *Journal of the Royal Statistical Society*. 1967; 29:513–524.
- Low Choy SL, O’Leary R, Mengersen K. Elicitation by design in ecology: Using expert opinion to inform priors for Bayesian statistical models. *Ecology*. 2009; 90:265–277. [PubMed: 19294931]
- MacKinnon DP. Analysis of mediating variables in prevention and intervention research. *NIDA research monograph*. 1994; 139:127–127. [PubMed: 8742555]
- MacKinnon, DP. *Introduction to statistical mediation analysis*. Taylor & Francis; 2008.
- MacKinnon DP, Fritz MS, Williams J, Lockwood CM. Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods*. 2007; 39:384–389. [PubMed: 17958149]
- MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*. 2002; 7:83–104. [PubMed: 11928892]
- MacKinnon DP, Lockwood CM, Williams J. Confidence limits for the indirect effect: distribution of the product and resampling methods. *Multivariate Behavioral Research*. 2004; 39:99–128. [PubMed: 20157642]

- MacKinnon, DP., Wurpts, IC., Valente, M. Imagery and memory as known effect validation of mediation analysis. Paper presented at the Society for Multivariate Experimental Psychology; Nashville, TN. 2014.
- McNeish DM. Using Data-Dependent Priors to Mitigate Small Sample Bias in Latent Growth Models: A Discussion and Illustration Using M plus. *Journal of Educational and Behavioral Statistics*. 2016a; 41(1):27–56.
- McNeish D. On Using Bayesian Methods to Address Small Sample Problems. *Structural Equation Modeling: A Multidisciplinary Journal*. 2016b; 23(5):750–773.
- Mio evi , M., MacKinnon, DP. SAS® for Bayesian Mediation Analysis. Proceedings of the 39th annual meeting of SAS Users Group International; Cary, NC: SAS Institute, Inc; 2014.
- Nuijten MB, Wetzels R, Matzke D, Dolan CV, Wagenmakers EJ. A default Bayesian hypothesis test for mediation. *Behavior Research Methods*. 2012:1–13. [PubMed: 21717266]
- O’Hagan A. Eliciting expert beliefs in substantial practical applications. *Journal of the Royal Statistical Society: Series D (The Statistician)*. 1998; 47(1):21–35.
- Preacher KJ, Hayes AF. SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, and Computers*. 2004; 36:717–731.
- SAS Institute Inc. SAS/STAT® 13.1 User’s Guide. Cary, NC: SAS Institute Inc; 2013.
- van de Schoot R, Hoijtink H, Mulder J, Van Aken MA, Orobio de Castro B, Meeus W, Romeijn JW. Evaluating expectations about negative emotional states of aggressive boys using Bayesian model selection. *Developmental psychology*. 2011; 47(1):203–212. [PubMed: 21244159]
- van de Schoot R, Kaplan D, Denissen J, Asendorpf JB, Neyer FJ, Aken MA. A Gentle Introduction to Bayesian Analysis: Applications to Developmental Research. *Child Development*. 2014; 85(3): 842–860. [PubMed: 24116396]
- Springer MD, Thompson WE. The distribution of independent random variables. *SIAM Journal on Applied Mathematics*. 1966; 14:511–526.
- Tofighi D, MacKinnon DP. RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods*. 2011; 43:692–700. [PubMed: 21487904]
- Valente MJ, Gonzalez O, Mio evi M, MacKinnon DP. 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*. 2015 0013164415618992.
- Yuan Y, MacKinnon DP. Bayesian mediation analysis. *Psychological Methods*. 2009; 14(4):301–322. [PubMed: 19968395]
- Wang L, Preacher KJ. Moderated mediation analysis using Bayesian methods. *Structural Equation Modeling: A Multidisciplinary Journal*. 2015; 22(2):249–263.
- Wilkinson L. Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*. 1999; 54(8):594–604.

Appendix

```

/
*****
*****/

/* The following code computes Bayesian credibility intervals for the
mediated effect */
/* and probabilities that the mediated effect is greater than 1, 2, and 5
from */
/* Mio evi , MacKinnon, and Levy,
2015. */
/
*****

```

```

*****/
* postint is a SAS autocall macro;
%include 'C:\Desktop\postint.sas';
/* The number of draws to burn-in (nbi), retain after burn-in (nmc), and the
thinning */
/* parameter were chosen based on diagnostic plots in SAS, and should be
changed by users */
/* to fit their models and data. The choices of distributional form and
hyperaparemeters */
/* for the prior distributions should also be modified by the user to fit
their specific */
/*
analysis.
*/

/* Bayesian analysis using diffuse priors for the method of
coefficients */
proc mcmc data=mydata outpost=outdiffcoef nbi=1000 nmc=5000 thin=5 seed=2;
parms a 0 b 0 cpr 0 i2 5 i3 5; *starting values for regression coefficients
and intercepts;
parms sigmaem 1 sigmaey 1; * starting values for residual variances of M and
Y;
prior a ~ normal (mean=0, var=1000); *prior for regression coefficient a;
prior b ~ normal (mean=0, var=1000);
prior cpr ~ normal (mean=0, var=1000);
prior i2 ~ uniform (1,9); *prior for the intercept in the equation
predicting M;
prior i3 ~ uniform (0,20);
prior sigmaem sigmaey ~ igamma(shape=0.001, scale=0.001); *priors for the
residual
variances of M and Y;
mum=int2+a*X; *mean function of M predicted by X;
muy=i3+b*m+cpr*X; *mean function of Y predicted by X and M;
model m~ n(mum, var=sigmaem); *model for M;
model y~ n(muy, var=sigmaey); *model for Y;
run;
/* Computation of the mediated effect from all retained
iterations */
data ab_coef; set outdiffcoef;
ab=a*b;
run;
/* Using the %postint macro to obtain equal-tail and HPD intervals for
ab */
%let parms = ab;
%postint(data=ab_coef, var=&parms, out=credint_coef);

```

```

/* Bayesian analysis using diffuse priors for the method of
covariances */
/* These diffuse priors reflect the assumption that X has a mean of .5 and
expected */
/* standard deviation of .15, that M has a mean of 5 and expected standard
deviation of 1, */
/* and that Y has a mean of 3 and expected standard deviation of 1. Another
assumption is */
/* that there is no covariance between X, M, and
Y. */
proc mcmc data=mydata outpost=outdiffcov nbi=1000 nmc=5000 thin=5 seed=2;
ods select PostSummaries PostIntervals;
    array data[3] X M Y; *specify variable names;
    array mu[3]; *specify the dimensions of the mean vector, i.e. number of
variables;
    array Sigma[3,3]; *specify the dimensions of the variance hyperparameter
for the mean vector;
    array mu0[3] (.5 5 3); *specify numerical values for the mean
hyperparameter in the
    multivariate normal prior;
    array Sigma0[3,3] (1000 0 0 0 1000 0 0 0 1000); *specify numerical values
for the variance
    hyperparameter in the multivariate normal prior;
    array SDIFFUSE[3,3] .0675 0 0 0 3 0 0 0 3; *specify numerical values for
S matrix in the
    inverse Wishart prior;
    parm mu Sigma;
    prior mu ~ mvn(mu0, Sigma0); *specify the hyperparameters for the
multivariate normal prior;
    prior Sigma ~ iwish(4, SDIFFUSE); *specify the hyperparameters for the
inverse Wishart prior;
    model data ~ mvn(mu, Sigma);
    run;
/* Computation of the mediated effect from all retained
iterations */
data ab_cov; set outdiffcov;
ab=(Sigma4/Sigma1)*(Sigma1*Sigma6-Sigma4*Sigma3)/(Sigma1*Sigma5-Sigma4**2);
RUN;
/* Using the %postint macro to obtain equal-tail and HPD intervals for
ab */
%let parms = ab;
%postint(data=ab_cov, var=&parms, out=credint_cov);
/* Bayesian analysis using informative priors for the method of
coefficients */

```



```

proc mcmc data=mydata outpost=outdiffcoefinfo nbi=1000 nmc=5000 thin=5
seed=2;
parms a 0 b 0 cpr 0 int2 0 i3 0;
parms sigmaem 1 sigmaey 1;
prior a ~ normal (mean=3.883, sd=0.769); *standard deviation parametrization
of the normal prior;
prior b ~ normal (mean=0.715, sd=0.206); * alternative parametrization are
var= for variance and
prec= for precision;
prior cpr ~ normal (mean=1.341, sd= 1.213);
prior int2 ~ normal (mean=4.250, sd=0.504);
prior i3 ~ normal (mean=8.112, sd=1.060);
prior sigmaem sigmaey ~ igamma(shape=0.001, scale=0.001);
mum=int2+a*X;
muy=i3+b*m+cpr*X;
model m~ n(mum, var=sigmaem);
model y~ n(muy, var=sigmaey);
run;
/* Computation of the mediated effect from all retained
iterations */
data ab_coef_info; set outdiffcoefinfo;
ab=a*b;
run;
/* Using the %postint macro to obtain equal-tail and HPD intervals for
ab */
%let parms = ab;
%postint(data=ab_coef_info, var=&parms, out=credint_coef_info);
/* Bayesian analysis using informative priors for the method of
covariances */
proc mcmc data=mydata outpost=outinfocov nbi=1000 nmc=5000 thin=5 seed=2;
ods select PostSummaries PostIntervals;
array data[3] X M Y;
array mu[3];
array Sigma[3,3];
array mu0[3] (0.429 5.914 12.914);
array Sigma0[3,3] (0.25210084 0.9789916 1.03781513 0.9789916 8.72773109
7.5512605
1.03781513 7.5512605 13.49243697);
array SDIFFUSE[3,3] 8.57143 33.28571 35.28571 33.28571 296.74286
256.74286 35.28571 256.74286 458.74286;
parm mu Sigma;
prior mu ~ mvn(mu0, Sigma0);
prior Sigma ~ iwish(35, SDIFFUSE);
model data ~ mvn(mu, Sigma);

```

```

run;
/* Computation of the mediated effect from all retained
iterations */
data ab_cov_info; set outinfocov;
ab=(Sigma4/Sigma1)*(Sigma1*Sigma6-Sigma4*Sigma3)/(Sigma1*Sigma5-Sigma4**2);
RUN;
/* Using the %postint macro to obtain equal-tail and HPD intervals for
ab */
%let parms = ab;
%postint(data=ab_cov_info, var=&parms, out=credint_cov_info);
/* print all of the credibilty intervals*/
title 'Credibility intervals for the method of coefficients with diffuse
priors';
proc print data=credint_coef;
run;
title 'Credibility intervals for the method of covariances with diffuse
priors';
proc print data=credint_cov;
run;
title 'Credibility intervals for the method of coefficients with informative
priors';
proc print data=credint_coef_info;
run;
title 'Credibility intervals for the method of covariances with informative
priors';
proc print data=credint_cov_info;
run;
/* Computing the probability of ab being equal to or greater than 1, 2, and
5*/
data ab_coef_diff; set ab_coef;
proppos11=0;
proppos12=0;
proppos15=0;
if ab = 1 then proppos11=1;
if ab gt 1 then proppos11=1;
if ab = 2 then proppos12=1;
if ab gt 2 then proppos12=1;
if ab = 5 then proppos15=1;
if ab gt 5 then proppos15=1;
run;
data ab_coef_info; set ab_coef_info;
proppos21=0;
proppos22=0;
proppos25=0;

```

```

if ab = 1 then proppos21=1;
if ab gt 1 then proppos21=1;
if ab = 2 then proppos22=1;
if ab gt 2 then proppos22=1;
if ab = 5 then proppos25=1;
if ab gt 5 then proppos25=1;
run;
data ab_cov_diff; set ab_cov;
proppos31=0;
proppos32=0;
proppos35=0;
if ab = 1 then proppos31=1;
if ab gt 1 then proppos31=1;
if ab = 2 then proppos32=1;
if ab gt 2 then proppos32=1;
if ab = 5 then proppos35=1;
if ab gt 5 then proppos35=1;
run;
data ab_cov_info; set ab_cov_info;
proppos41=0;
proppos42=0;
proppos45=0;
if ab = 1 then proppos41=1;
if ab gt 1 then proppos41=1;
if ab = 2 then proppos42=1;
if ab gt 2 then proppos42=1;
if ab = 5 then proppos45=1;
if ab gt 5 then proppos45=1;
run;
title 'Probabilities for the method of coefficients with diffuse priors';
proc means data=ab_coef_diff;
var proppos11 proppos12 proppos15;
run;
title 'Probabilities for the method of coefficients with informative priors';
proc means data=ab_coef_info;
var proppos21 proppos22 proppos25;
run;
title 'Probabilities for the method of covariances with diffuse priors';
proc means data=ab_cov_diff;
var proppos31 proppos32 proppos35;
run;
title 'Probabilities for the method of covariances with informative priors';
proc means data=ab_cov_info;

```

```
var proppos41 proppos42 proppos45;  
run;
```

Figure 1.
Single mediator model.

Figure 2.

Trellis plot of Power for all methods and all parameter combinations as a function of sample size in Study 1. The letter markers indicate the following: D (distribution) codes normal theory (gray solid line) and distribution of the product (black solid line) confidence limits, Y (Yuan & MacKinnon, 2009) codes credibility intervals formed using the Bayesian method of coefficients with diffuse (gray dashed line) and informative (black dashed line) prior distributions, E (Enders, Fairchild, and MacKinnon, 2013) codes credibility intervals formed using the Bayesian method of covariances with diffuse (gray dotted line) and informative (black dotted line) distributions, and B (bootstrap) codes the percentile (gray dot-dash line) and bias-corrected (black dot-dash line) bootstrap confidence limits.

Figure 3.
Plots of Power as a function of precision in the prior for the regression coefficient b in Study 2.

Table 1

Percentages of true precision values of a and b represented by the values of 10 raised to the powers of -2, -1, 0, 1, and 2 used as precision parameters in prior distributions for a and b.

N	prec(a)					prec(b)				
	10 ⁻²	10 ⁻¹	10 ⁰	10 ¹	10 ²	10 ⁻²	10 ⁻¹	10 ⁰	10 ¹	10 ²
20	0.10	0.60	5.60	55.60	555.60	0.00	0.60	5.90	58.80	588.20
40	0.00	0.30	2.60	26.30	263.20	0.00	0.30	2.70	27.00	270.30
60	0.00	0.20	1.70	17.20	172.40	0.00	0.20	1.70	17.50	175.40
100	0.00	0.10	1.00	10.20	102.00	0.00	0.10	1.00	10.30	103.10
200	0.00	0.00	0.50	5.00	50.50	0.00	0.00	0.50	5.10	50.80

Table 2

Prior specifications for Bayesian analyses in the Empirical Example

Method	Verbal description of priors	Notation
Bayesian method of coefficients with diffuse priors	Diffuse normal prior for regression coefficients with the mean equal to zero and the variance parameter equal to 1000, uniform priors for intercepts with limits equal to the ranges of M and Y, and diffuse inverse gamma priors with both the scale and shape parameters equal to 0.001 for residual variances	$a, b, c' \sim N(0, 1000)$ $i_{\mathcal{M}} \sim \text{Unif}(1, 9)$ $i_{\mathcal{Y}} \sim \text{Unif}(0, 20)$ $\sigma_{\text{eM}}^2, \sigma_{\text{eY}}^2 \sim \text{IG}(0.001, 0.001)$
Bayesian method of covariances with diffuse priors	Diffuse multivariate normal prior with a vector of zeros as the mean parameter and a diagonal matrix with each diagonal element equal to 1000 for the mean vector and an inverse-Wishart prior with 4 degrees of freedom and a diagonal matrix with each diagonal element equal to 3 for the covariance matrix of X, M, and Y	$\underline{\mu} \sim \text{MVN}\left(\begin{bmatrix} 0.5 \\ 5 \\ 3 \end{bmatrix}, \begin{bmatrix} 1000 & 0 & 0 \\ 0 & 1000 & 0 \\ 0 & 0 & 1000 \end{bmatrix}\right)$ $\underline{\Sigma} \sim \text{IW}(4, \begin{bmatrix} 0.0675 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix})$
Bayesian method of coefficients with informative priors	Informative normal prior for regression coefficients and intercepts with the mean parameter equal to the observed regression coefficient (or intercept) in the prior data set and the standard deviation parameter equal to the standard error of the corresponding regression coefficient and diffuse inverse gamma priors with both the scale and shape parameters equal to 0.001 for residual variances	$a \sim N(3.883, 0.769)$ $b \sim N(0.715, 0.206)$ $c' \sim N(1.341, 1.213)$ $i_2 \sim N(4.250, 0.504)$ $i_3 \sim N(8.112, 1.060)$ $\sigma_{\text{eM}}^2, \sigma_{\text{eY}}^2 \sim \text{IG}(0.001, 0.001)$
Bayesian method of covariances with informative priors	Informative multivariate normal prior with a vector of observed means from the prior sample as the mean parameter and the observed covariance matrix as the variance parameter for the mean vector and an inverse-Wishart prior with degrees of freedom equal to the sample size of the prior sample and the observed covariation matrix as the positive definite matrix S for the covariance matrix of X, M, and Y	$\underline{\mu} \sim \text{MVN}\left(\begin{bmatrix} 0.429 \\ 5.914 \\ 12.914 \end{bmatrix}, \begin{bmatrix} 0.252 & 0.979 & 1.038 \\ 0.979 & 8.728 & 7.551 \\ 1.038 & 7.551 & 13.492 \end{bmatrix}\right)$ $\underline{\Sigma} \sim \text{IW}(35, \begin{bmatrix} 8.571 & 33.286 & 35.286 \\ 33.286 & 296.743 & 256.743 \\ 35.286 & 256.743 & 458.743 \end{bmatrix})$

Table 3

Results of the Empirical Example

Method	Confidence interval	Equal-tail credibility interval	HPD credibility interval	P(ab 1)	P(ab 2)	P(ab 5)
Normal theory	[0.378, 3.991]	-	-	-	-	-
Distribution of the product	[0.560, 4.169]					
Percentile bootstrap	[0.769, 3.583]	-	-	-	-	-
Bias-corrected bootstrap	[0.870, 3.614]	-	-	-	-	-
Bayesian method of coefficients with diffuse priors	-	[0.447, 4.017]	[0.364, 3.894]	0.910	0.563	0.002
Bayesian method of covariances with diffuse priors	-	[0.629, 3.983]	[0.530, 3.845]	0.917	0.557	0.003
Bayesian method of coefficients with informative priors	-	[1.487, 3.786]	[1.494, 3.787]	0.995	0.811	0.000
Bayesian method of covariances with informative priors	-	[1.304, 3.822]	[1.302, 3.799]	0.993	0.752	0.000

Table 4

Results of the Sensitivity Analysis

Method	Prior	95% CI for ab equal-tail HPD
Bayesian method of coefficients with diffuse priors	baseline informative prior $a \sim N(3.883, 0.769)$ $b \sim N(0.715, 0.206)$ $c' \sim N(1.341, 1.213)$ $i_2 \sim N(4.250, 0.504)$ $i_3 \sim N(8.112, 1.060)$ $\sigma^2_{eM}, \sigma^2_{eY} \sim IG(0.001, 0.001)$	[1.487, 3.786] [1.494, 3.787]
	two times more informative $a \sim N(3.883, 0.385)$ $b \sim N(0.715, 0.103)$ $c' \sim N(1.341, 0.607)$ $i_2 \sim N(4.250, 0.252)$ $i_3 \sim N(8.112, 0.530)$	[1.985, 3.413] [2.020, 3.422]
	two times less informative $a \sim N(3.883, 1.538)$ $b \sim N(0.715, 0.412)$ $c' \sim N(1.341, 2.426)$ $i_2 \sim N(4.250, 1.008)$ $i_3 \sim N(8.112, 2.120)$	[0.970, 3.972] [0.887, 3.777]
	five times less informative $a \sim N(3.883, 3.845)$ $b \sim N(0.715, 1.030)$ $c' \sim N(1.341, 6.065)$ $i_2 \sim N(4.250, 2.520)$ $i_3 \sim N(8.112, 5.300)$	[0.647, 4.360] [0.700, 4.379]
	ten times less informative $a \sim N(3.883, 7.690)$ $b \sim N(0.715, 2.060)$ $c' \sim N(1.341, 12.130)$ $i_2 \sim N(4.250, 5.040)$ $i_3 \sim N(8.112, 10.600)$	[0.539, 4.083] [0.516, 3.977]
Bayesian method of covariances with informative priors	baseline informative prior $\underline{\mu} \sim MVN\left(\begin{bmatrix} 0.429 \\ 5.914 \\ 12.914 \\ 8.571 \end{bmatrix}, \begin{bmatrix} 0.252 & 0.979 & 1.038 \\ 0.979 & 8.728 & 7.551 \\ 1.038 & 7.551 & 13.492 \\ 8.571 & 33.286 & 35.286 \end{bmatrix}\right)$ $\Sigma \sim IW(35, \begin{bmatrix} 33.286 & 296.743 & 256.743 \\ 35.286 & 256.743 & 458.743 \end{bmatrix})$	[1.304, 3.822] [1.302, 3.799]
	two times more informative $\Sigma \sim IW(70, \begin{bmatrix} 8.571 & 33.286 & 35.286 \\ 33.286 & 296.743 & 256.743 \\ 35.286 & 256.743 & 458.743 \end{bmatrix})$	[1.443, 3.628] [1.498, 3.647]
	two times less informative $\Sigma \sim IW(17.5, \begin{bmatrix} 8.571 & 33.286 & 35.286 \\ 33.286 & 296.743 & 256.743 \\ 35.286 & 256.743 & 458.743 \end{bmatrix})$	[1.034, 4.048] [0.961, 3.899]
	five times less informative $\Sigma \sim IW(7, \begin{bmatrix} 8.571 & 33.286 & 35.286 \\ 33.286 & 296.743 & 256.743 \\ 35.286 & 256.743 & 458.743 \end{bmatrix})$	[0.901, 4.263] [0.852, 4.137]
	ten times less informative $\Sigma \sim IW(3.5, \begin{bmatrix} 8.571 & 33.286 & 35.286 \\ 33.286 & 296.743 & 256.743 \\ 35.286 & 256.743 & 458.743 \end{bmatrix})$	[0.896, 4.241] [0.893, 4.226]

* Note: In the method of coefficients the normal priors are specified using the mean and standard deviation parametrization and the prior for the residual variances of M and Y remained the same for all five situations, thus it was not repeated in the table. In the method of covariances the prior for the mean vector did not change between the four cases, thus it was not repeated in the table.