A default Bayesian hypothesis test for mediation

Nuijten, M.B.; Wetzels, R.; Matzke, D.; Dolan, C.V.; Wagenmakers, E.-J.

DOI
10.3758/s13428-014-0470-2

Publication date
2015

Document Version
Final published version

Published in
Behavior Research Methods

Behav Res (2015) 47:85–97
DOI 10.3758/s13428-014-0470-2

A default Bayesian hypothesis test for mediation

Michèle B. Nuijten · Ruud Wetzels · Dora Matzke ·
Conor V. Dolan · Eric-Jan Wagenmakers

Published online: 6 June 2014

Abstract In order to quantify the relationship between multiple variables, researchers often carry out a mediation analysis. In such an analysis, a mediator (e.g., knowledge of a healthy diet) transmits the effect from an independent variable (e.g., classroom instruction on a healthy diet) to a dependent variable (e.g., consumption of fruits and vegetables). Almost all mediation analyses in psychology use frequentist estimation and hypothesis-testing techniques. A recent exception is Yuan and MacKinnon (*Psychological Methods, 14,* 301–322, 2009), who outlined a Bayesian parameter estimation procedure for mediation analysis. Here we complete the Bayesian alternative to frequentist mediation analysis by specifying a default Bayesian hypothesis test based on the Jeffreys–Zellner–Siow approach. We further extend this default Bayesian test by allowing a comparison to directional or one-sided alternatives, using Markov chain Monte Carlo techniques implemented in JAGS. All Bayesian tests are implemented in the R package BayesMed (Nuijten, Wetzels, Matzke, Dolan, & Wagenmakers, 2014).

Keywords Bayes factor · Evidence · Mediated effects

Electronic supplementary material The online version of this article (doi:10.3758/s13428-014-0470-2) contains supplementary material, which is available to authorized users.

M. B. Nuijten
Tilburg University, Tilburg, The Netherlands

R. Wetzels · D. Matzke · E.-J. Wagenmakers (✉)
Department of Psychology, University of Amsterdam, Weesperplein 4, 1018 XA Amsterdam, The Netherlands
e-mail: EJ.Wagenmakers@gmail.com

C. V. Dolan
VU University Amsterdam, Amsterdam, The Netherlands

Mediated relationships are central to the theory and practice of psychology. In the prototypical scenario, a mediator (*M*; e.g., knowledge of a healthy diet) transmits the effect from an independent variable (*X*; e.g., classroom instruction on a healthy diet) to a dependent variable (*Y*; e.g., consumption of fruits and vegetables). Other examples arise in social psychology, where attitudes (*X*) cause intentions (*M*) and these intentions affect behavior (*Y*; MacKinnon, Fairchild, & Fritz, 2007). To quantify such relationships between mediator, independent variable, and dependent variable, researchers often use a toolbox of popular statistical methods collectively known as mediation analysis.

The currently available tools for mediation analyses are almost exclusively based on classical or frequentist statistics, featuring concepts such as confidence intervals and *p* values. Recently, Yuan and MacKinnon (2009) proposed a Bayesian mediation analysis that allows researchers to obtain a posterior distribution (and associated credible interval) for the mediated effect. This posterior distribution quantifies the uncertainty about the strength of the mediated effect under the assumption that the effect does not equal zero. This approach constitutes a valuable addition to the toolbox of mediation methods, but it specifically concerns parameter estimation and not hypothesis testing. As Yuan and MacKinnon stated in their conclusion: "One important topic we have not covered in this article is hypothesis testing . . . Strict Bayesian hypothesis testing is based on Bayes factor, which is essentially the odds of the null hypothesis being true versus the alternative hypothesis being true, conditional on the observed data. The use of Bayesian hypothesis testing . . . would be a reasonable future research topic in Bayesian mediation analysis."

Hence, the goal of this article is to add another statistical method to the toolbox of mediation analysis—namely, the Bayes factor hypothesis test alluded to by Yuan and MacKinnon (2009). In the development of this test, we have assumed a default specification of prior distributions based on the Jeffreys–Zellner–Siow framework (Liang, Paulo, Molina, Clyde, & Berger, 2008), promoted in psychology by Jeff Rouder, Richard Morey, and colleagues (Rouder & Morey, 2012; Rouder, Morey, Speckman, & Province, 2012; Rouder, Speckman, Sun, Morey, & Iverson, 2009), as well as ourselves (Wetzels, Grasman, & Wagenmakers, 2012; Wetzels, Raaijmakers, Jakab, & Wagenmakers, 2009; Wetzels & Wagenmakers, 2012; for an alternative approach, see Semmens-Wheeler, Dienes, & Duka, 2013). In our opinion, the default specification of prior distributions is useful because it provides a reference analysis that can be carried out regardless of subjective considerations about the topic at hand. Of course, researchers who have prior knowledge may wish to incorporate that knowledge into the models to devise a more informative test (e.g., Armstrong & Dienes, 2013; Dienes, 2011; Guo, Li, Yang, & Dienes, 2013). Here, we focus solely on the default test as it pertains to the prototypical, single-level scenario of three variables.

The outline of this article is as follows. First, we briefly discuss the conventional frequentist tests and the existing Bayesian mediation analysis proposed by Yuan and MacKinnon (2009). We then explain Bayesian hypothesis testing in general and introduce our default Bayesian hypothesis test for mediation. We illustrate the performance of our test with a simulation study and an example of a psychological study. Finally, we discuss software in which we implemented the Bayesian methods for mediation analysis: the R package BayesMed (Nuijten et al., 2014).

Frequentist mediation analysis

Consider a relation between an independent variable X and a dependent variable Y (see Fig. 1a). In a linear regression equation, such a relation can be represented as follows:

$$Y_i = \beta_{0(1)} + \tau X_i + \varepsilon_{(1)}, \tag{1}$$

where subscript i identifies the participant, τ represents the relation between the independent variable X and the dependent variable Y, $\beta_{0(1)}$ is the intercept, and $\varepsilon_{(1)}$ is the residual. The effect of X on Y, path τ, is called the *total effect*.

The relation between X and Y can be mediated by variable M, which means that a change in X leads to a change in M, which then leads to a change in Y (see Fig. 1b, c). The resulting mediation model can be represented by the following set of linear regression equations:

$$Y_i = \beta_{0(2)} + \tau' X_i + \beta M_i + \varepsilon_{(2)}, \tag{2}$$

$$M_i = \beta_{0(3)} + \alpha X_i + \varepsilon_{(3)}, \tag{3}$$

where τ' represents the relation between X and Y after adjusting for the effects of the mediator M, α represents the relation between X and M, and β represents the relation between M and Y. Furthermore, $\varepsilon_{(1)}$, $\varepsilon_{(2)}$, and $\varepsilon_{(3)}$ are assumed to be conditionally normally distributed, independent, homoskedastic residuals. Throughout the remainder of this article, we focus on the standardized mediation model (i.e., a model in which the variables are standardized) and refer to the regression coefficients α, β, and τ' as *paths*.

The product of α and β is the indirect effect, or the mediated effect, assuming that α and β are independent. The remaining direct effect of X on Y is denoted with τ'. If the mediated effect differs from zero and τ' equals zero, the effect of X on Y is completely mediated by M (see Fig. 1c). If τ' has a value other than zero, the relationship between X and Y is only partially mediated by M (see Fig. 1b).

A popular method to test for mediation is to test paths α and β simultaneously. The estimated indirect effect $\widehat{\alpha}\widehat{\beta}$ is divided by its standard error, and the resulting Z statistic is compared with the standard normal distribution to assess whether the effect is significantly different from zero—in which case, the null hypothesis of no mediation can be rejected.

There exist several ways to calculate the standard error of $\widehat{\alpha}\widehat{\beta}$, but the one used in the Sobel test (Sobel, 1982) is commonly reported:

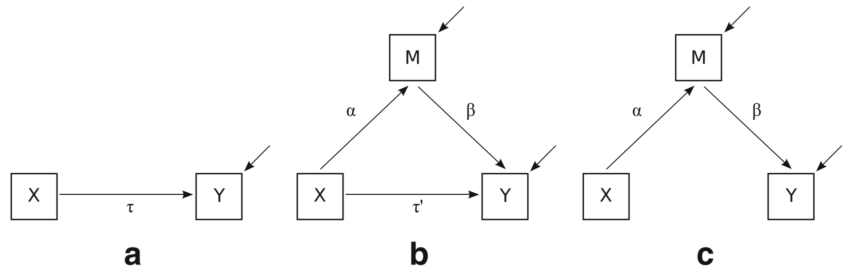
$$\widehat{\sigma}_{\widehat{\alpha}\widehat{\beta}} = \sqrt{\widehat{\beta}^2 \widehat{\sigma}^2_\alpha + \widehat{\alpha}^2 \widehat{\sigma}^2_\beta}, \tag{4}$$

where $\widehat{\alpha}$ and $\widehat{\beta}$ are the point estimates of the regression coefficients of the mediated effect and $\widehat{\sigma}_a$ and $\widehat{\sigma}_\beta$ their standard errors. The 95 % confidence interval for the mediated effect is then given by $\widehat{\alpha}\widehat{\beta} \pm 1.96 \times \widehat{\sigma}_{\widehat{\alpha}\widehat{\beta}}$.

One problem with the Sobel test is that it assumes a symmetrical sampling distribution for the mediated effect, whereas in reality this distribution is skewed (MacKinnon, Lockwood, & Hoffman, 1998). Consequently, the Sobel test has relatively low power (MacKinnon, Warsi, & Dwyer, 1995). A solution to this problem is to construct a confidence interval that takes the asymmetry of the distribution into account (see, e.g., the product method of MacKinnon, Lockwood, Hoffman, & West, 2002) or the profile likelihood method (see Venzon & Moolgavkar, 1988).

Our goal here is not to argue against frequentist statistics in general, or p values in particular; for this, we refer the interested reader to the following articles and references therein: Berger and Delampady (1987), Berger and Wolpert (1988), Dienes (2011), Edwards, Lindman, and Savage (1963), O'Hagan and Forster (2004), Rouder et al. (2012), Sellke, Bayarri, and Berger (2001), Wagenmakers (2007); Wetzels et al., (2011). Instead, our goal is to outline an additional Bayesian tool that can be used for mediation analysis. The availability of multiple tools is useful, not just because different situations may require different tools, but also because they allow a robustness check; if different tools yield opposing conclusions, the careful researcher does well to report the results from both tests, indicating that the data are ambiguous

Fig. 1 Diagram of the standard mediation model. **a** Direct relation between X and Y, panel. **b** Partial mediation. **c** Full mediation. Diagonal arrows indicate that the graphical node is perturbed by an error term



in the sense that the conclusion depends on the analysis method at hand.

An alternative: Bayesian estimation

Our end goal is to propose a Bayesian alternative for the frequentist mediation test. Below, we consider the Bayesian treatment of the mediation model in detail, but first we briefly discuss Bayesian inference in general terms. In the Bayesian framework, uncertainty is quantified by probability. Prior beliefs about parameters are formalized by prior probability distributions that are updated by the observed data to result in posterior beliefs or posterior distributions (Dienes, 2008; Kruschke, 2010; Lee & Wagenmakers, 2013; O'Hagan & Forster, 2004).

The Bayesian updating process proceeds as follows. First, before observing the data under consideration, the Bayesian statistician assigns a probability distribution to one or more model parameters θ on the basis of his or her prior knowledge; hence, this distribution is known as the prior probability distribution, or simply the *prior*, denoted $p(\theta)$. Next one observes data $\mathbf{D}$, and the statistical model can be used to calculate the associated probability of $\mathbf{D}$ occurring under specific values of θ, a quantity known as the likelihood, denoted $p(\mathbf{D}|\theta)$. The prior distribution $p(\theta)$ is then updated to the posterior distribution $p(\theta|\mathbf{D})$ according to Bayes' rule:

$$p(\theta|\mathbf{D}) = \frac{p(\mathbf{D}|\theta)p(\theta)}{p(\mathbf{D})}. \quad (5)$$

Note that the marginal likelihood $p(\mathbf{D})=\int p(\mathbf{D}|\theta)p(\theta)\, d\theta$ functions as a normalizing constant that ensures that the posterior distribution will integrate to one. Because the normalizing constant does not contain θ, it is not important for parameter estimation, and Eq. 5 is often written as follows:

$$p(\theta|\mathbf{D}) \propto p(\mathbf{D}|\theta)p(\theta), \quad (6)$$

or in words:

$$\text{Posterior Distribution} \propto \text{Likelihood} \times \text{Prior Distribution}, \quad (7)$$

where $\propto$ means *proportional to*.

In a Bayesian mediation analysis, the above updating principle can be used to transition from prior to posterior distributions for parameters α, β, and τ', as proposed by Yuan and MacKinnon (2009). Their method allows the user to determine the posterior distribution of the indirect effect $\alpha\beta$, together with a 95 % credible interval. This interval has the intuitive interpretation that we can be 95 % confident that the true value of $\alpha\beta$ resides within this interval.

The approach of Yuan and MacKinnon (2009) is appropriate when estimating the size of the mediated effect. However, in experimental psychology, the research question is often framed in terms of model selection or hypothesis testing; that is, the researcher seeks to answer the question: "does the effect exist?" Parameter estimation and model selection have different aims and, depending on the situation at hand, one procedure may be more appropriate than the other. We contend that there are situations where a hypothesis test is scientifically useful (e.g., Iverson, Wagenmakers, & Lee, 2010; Rouder et al., 2009), and in what follows, we proceed to outline a default Bayesian hypothesis test for mediation. In order to keep this article self-contained, we will first introduce the principles of Bayesian hypothesis testing (Hoijtink, Klugkist, & Boelen, 2008; Myung & Pitt, 1997; Vandekerckhove, Matzke, & Wagenmakers, in press; Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010).

Bayesian hypothesis testing

A Bayesian hypothesis test is a model selection procedure with two models or hypotheses. Assume two competing models or hypotheses, $\mathcal{M}_0$ and $\mathcal{M}_1$, with respective a priori plausibility $p(\mathcal{M}_0)$ and $p(\mathcal{M}_1)=1-p(\mathcal{M}_0)$. Differences in prior plausibility are often subjective but can be used to formalize the idea that extraordinary claims require extraordinary evidence (Lee & Wagenmakers, 2013, Chap. 7). The ratio $p(\mathcal{M}_1)/p(\mathcal{M}_0)$ is known as the prior model odds. The data update the prior model odds to arrive at the posterior model odds, $p(\mathcal{M}_1|\mathbf{D})/p(\mathcal{M}_0|\mathbf{D})$, as follows:

$$\frac{p(\mathcal{M}_1|\mathbf{D})}{p(\mathcal{M}_0|\mathbf{D})} = \frac{p(\mathbf{D}|\mathcal{M}_1)p(\mathcal{M}_1)}{p(\mathbf{D}|\mathcal{M}_0)p(\mathcal{M}_0)}. \quad (8)$$

or in words:

$$\text{Posterior Model Odds} = \text{Bayes Factor} \times \text{Prior Model Odds}. \quad (9)$$

Equation 8 shows that the change in model odds brought about by the data is given by the so-called Bayes factor (Jeffreys, 1961) which is the ratio of marginal likelihoods (i.e., normalizing constants in Eq. 5):

$$BF_{10} = \frac{p(\mathbf{D}|\mathcal{M}_1)}{p\left(\mathbf{D}\middle|\mathcal{M}_0\right)}. \tag{10}$$

The Bayes factor quantifies the weight of evidence for $\mathcal{M}_1$ versus $\mathcal{M}_0$ that is provided by the data and, as such, it represents "the standard Bayesian solution to the hypothesis testing and model selection problems" (Lewis & Raffery, 1997, p. 648) and "the primary tool used in Bayesian inference for hypothesis testing and model selection" (Berger, 2006, p. 378).

When $BF_{10}>1$, this indicates that the data are more likely under $\mathcal{M}_1$, and when $BF_{10}<1$, this indicates that the data are more likely under $\mathcal{M}_0$. For example, when $BF_{10}=0.08$, the observed data are 12.5 times more likely under $\mathcal{M}_0$ than under $\mathcal{M}_1$ (i.e., $BF_{01}=1/BF_{10}=1/.08=12.5$). Note that the Bayes factor allows researchers to quantify evidence in favor of the null hypothesis.

Even though the default Bayes factor has an unambiguous and continuous scale, it is sometimes useful to summarize the Bayes factor in terms of discrete categories of evidential strength. Jeffreys (1961, Appendix B) proposed the classification scheme shown in Table 1. We replaced the labels "worth no more than a bare mention" with "anecdotal," "decisive" with "extreme," and "substantial" with "moderate." These labels facilitate scientific communication but should be considered only as an approximate descriptive articulation of different standards of evidence.

Under equal prior odds, Bayes factors can be converted to posterior probabilities $p(\mathcal{M}_1|\mathbf{D})=BF_{10}/(BF_{10}+1)$. This means that, for example, $BF_{10}=2$ translates to $p(\mathcal{M}_1|\mathbf{D})=2/3$.

Bayesian hypothesis test for mediation

The Bayesian hypothesis test for mediation contrasts the following two models:

$$\begin{aligned}\mathcal{M}_0 &: \alpha\beta = 0,\\ \mathcal{M}_1 &: \alpha\beta \neq 0.\end{aligned} \tag{11}$$

Observe that $\mathcal{M}_1$ entails that both $\alpha\neq0$ and $\beta\neq0$, so that BF_{10} can be obtained by combining the evidence for the presence of the two paths. Furthermore, note that in the standardized model, path α equals the correlation r_{XM} and path β equals the partial correlation $r_{MY|X}$. This means that we can use the existing default Bayesian hypothesis tests for correlation and partial correlation (Wetzels & Wagenmakers, 2012) and combine the evidence for the presence of the separate paths to yield the overall Bayes factor for mediation.

Table 1 Evidence categories for the Bayes factor BF_{10} (Jeffreys, 1961)

Bayes Factor BF_{10}	Interpretation
>100	Extreme evidence for $\mathcal{M}_1$
30–100	Very Strong evidence for $\mathcal{M}_1$
10–30	Strong evidence for $\mathcal{M}_1$
3–10	Moderate evidence for $\mathcal{M}_1$
1–3	Anecdotal evidence for $\mathcal{M}_1$
1	No evidence
1/3–1	Anecdotal evidence for $\mathcal{M}_0$
1/10–1/3	Moderate evidence for $\mathcal{M}_0$
1/30–1/10	Strong evidence for $\mathcal{M}_0$
1/100–1/30	Very Strong evidence for $\mathcal{M}_0$
<1/100	Extreme evidence for $\mathcal{M}_0$

Note. We replaced the labels "not worth more than a bare mention" with "anecdotal," "decisive" with "extreme," and "substantial" with "moderate."

The default JZS prior

The construction of good default priors is an active area of research in Bayesian statistics (e.g., Consonni, Forster, & La Rocca, 2013; Overstall & Forster, 2010). Most work in this area has been done in the context of linear regression. It is therefore advantageous to formulate the tests for correlation and partial correlation in terms of linear regression, so that existing developments for the selection of default priors can be brought to bear.

A popular default prior for linear regression is Zellner's g prior, which includes a normal distribution on the regression coefficients $\boldsymbol{\alpha}$, Jeffreys's (1961) prior on the precision ϕ (i.e., a prior that is invariant under transformation), and a uniform prior on the intercept β_0:

$$\begin{aligned} p(\boldsymbol{\alpha}|\phi, g, \mathbf{X}) &\sim N\left(0, \frac{g}{\phi}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right),\\ p(\phi) &\propto \frac{1}{\phi},\\ p(\beta_0) &\propto 1,\end{aligned} \tag{12}$$

where $\mathbf{X}$ denotes the matrix of predictor variables and the precision ϕ is the inverse of the variance. The coefficient g is a scaling factor and controls the weight of the prior relative to the weight of the data. For example, if $g=1$, the prior has exactly as much weight as the data, and if $g=10$, the prior has one tenth of the weight of the data. A popular default choice is $g=n$, the unit information prior, where the prior has as much influence as a single observation (Kass & Wasserman, 1995) and the behavior of the test becomes similar to that of BIC (Schwarz, 1978).

However, Liang et al. (2008) showed that the above specification yields a bound on the Bayes factor, even when there is overwhelming information supporting $\mathcal{M}_1$. This "information paradox" can be overcome by assigning the regression coefficients a Cauchy prior instead of a normal prior (Zellner & Siow, 1980). Equivalently, this can be accomplished by assigning g from Eq. 12 an Inverse-Gamma(1/2,n/2) prior:

$$p(\boldsymbol{\alpha}|\phi, g, \mathbf{X}) \sim N\left(0, \frac{g}{\phi}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right),$$
$$p(g) = \frac{(n/2)^{1/2}}{\Gamma(1/2)} g^{(-3/2)} e^{-n/(2g)}, \quad (13)$$
$$p(\phi) \propto \frac{1}{\phi}.$$

The above specification is known as the Jeffreys–Zellner–Siow, or JZS, prior. The JZS prior was adopted by Wetzels and Wagenmakers (2012) for the default tests of correlation and partial correlation, and the same tests are used here to compute the Bayes factor for mediation. It should be stressed, however, that the framework is general and allows researchers to add substantive knowledge about the topic under study by changing the prior distributions (e.g., Armstrong & Dienes, 2013; Dienes, 2011; Guo et al., 2013).

With the JZS tests for correlation and partial correlation in hand, we created the default Bayesian hypothesis test for mediation in three steps, as described in the next paragraphs.

Step 1: evidence for path α

The first step in the hypothesis test for mediation is to establish the Bayes factor for a correlation between X and M, path α (see Fig. 1). This test can be formulated as a comparison between two linear models:

$$\mathcal{M}_0 : M = \beta_0 + \varepsilon,$$
$$\mathcal{M}_1 : M = \beta_0 + \alpha X + \varepsilon, \quad (14)$$

where ε is the normally distributed error term. The default JZS Bayes factor quantifies the extent to which the data support $\mathcal{M}_1$ with path α versus $\mathcal{M}_0$ without path α, as follows (Wetzels & Wagenmakers, 2012):

$$\begin{aligned} BF_{10} &= BF_{\alpha} \\ &= \frac{P(\mathbf{D}|\mathcal{M}_1)}{P(\mathbf{D}|\mathcal{M}_0)} \\ &= \frac{(n/2)^{1/2}}{\Gamma(1/2)} \times \int_0^{\infty} (1+g)^{(n-2)/2} \times \left[1 + (1-r^2)g\right]^{-(n-1)/2} g^{(-3/2)} e^{-n/(2g)} \, dg, \end{aligned} \quad (15)$$

where n is the number of observations and r is the sample correlation.

For the proposed mediation test, we have to multiply the posterior probabilities of paths α and β, since both independent paths need to be present for mediation to hold. Hence, we need to convert the Bayes factor for path α to a posterior probability. Under the assumption of equal prior odds, this conversion is straightforward:

$$p(\alpha \neq 0|\mathbf{D}) = \frac{BF_{\alpha}}{BF_{\alpha} + 1}. \quad (16)$$

Step 2: evidence for path β

The second step in the hypothesis test for mediation is to establish the Bayes factor for a unique correlation between M and Y (without any influence from X), path β (see Fig. 1). Again, this test can be formulated as a comparison between two linear models:

$$\mathcal{M}_0 : Y = \beta_0 + \tau X + \varepsilon,$$
$$\mathcal{M}_1 : Y = \beta_0 + \tau' X + \beta M + \varepsilon, \quad (17)$$

where ε is the normally distributed error term. The default JZS Bayes factor quantifies the extent to which the data support $\mathcal{M}_1$ with path β versus $\mathcal{M}_0$ without path β, as in a test for partial correlation (Wetzels & Wagenmakers, 2012):

$$\begin{aligned} BF_{10} &= BF_{\beta} \\ &= \frac{P(\mathbf{D}|\mathcal{M}_1)}{P\left(\mathbf{D}\middle|\mathcal{M}_0\right)} \\ &= \frac{\int_0^{\infty} (1+g)^{(n-1-p_1)/2} \times \left[1 + (1-r_1^2)g\right]^{-(n-1)/2} g^{(-3/2)} e^{-n/(2g)} \, dg}{\int_0^{\infty} (1+g)^{(n-1-p_0)/2} \times \left[1 + (1-r_0^2)g\right]^{-(n-1)/2} g^{(-3/2)} e^{-n/(2g)} \, dg}, \end{aligned} \quad (18)$$

where n is the number of observations, r_1^2 and r_0^2 represent the explained variance of $\mathcal{M}_1$ and $\mathcal{M}_0$, respectively, and $p_1=2$ and $p_0=1$ are the number of regression coefficients or paths in $\mathcal{M}_1$ and $\mathcal{M}_0$, respectively.

As before, we can convert the Bayes factor for β to a posterior probability under the assumption of equal prior odds:

$$p(\beta \neq 0|\mathbf{D}) = \frac{BF_{\beta}}{BF_{\beta} + 1}. \quad (19)$$

Step 3: evidence for mediation

The third step in the hypothesis test for mediation is to multiply the evidence for α with the evidence for β to obtain the overall evidence for mediation:

$$\text{Evidence for Mediation} = p(\alpha \neq 0|\mathbf{D}) \times p(\beta \neq 0|\mathbf{D}). \quad (20)$$

The resulting evidence for mediation is a posterior probability that ranges from zero when there is no evidence for mediation at all to one when there is absolute certainty that mediation is present. We can also express the evidence for mediation as a Bayes factor through a simple transformation:

$$BF_{med} = \frac{\text{Evidence for Mediation}}{1-\text{Evidence for Mediation}}, \quad (21)$$

where a $BF_{med}>1$ indicates evidence for mediation and $BF_{med} <1$ indicates evidence against mediation.

Note that we can multiply the posterior probabilities, because the estimates of path α and β are uncorrelated. This can be demonstrated by inspecting the relevant element of the inverse of the information matrix—that is, the matrix of second order derivatives of the parameters α and β, with respect to the log likelihood function. This can be done numerically, since most SEM programs supply this matrix, and it can be done analytically. These results can be found in the supplemental materials.

Testing for full or partial mediation

An optional fourth step in the hypothesis test for mediation is to assess the evidence for full versus partial mediation. The relation between X and Y is fully mediated by M when $\alpha\beta$ differs from zero and the direct path between X and Y, path τ' is zero. The evidence for τ' can be assessed with the JZS test for partial correlation as we did for path β (see Eq. 18). Note, however, that the specification of the null model has changed:

$$\begin{aligned} \mathscr{M}_0 &: Y = \beta_0 + \beta M + \varepsilon, \\ \mathscr{M}_1 &: Y = \beta_0 + \tau' X + \beta M + \varepsilon. \end{aligned} \tag{22}$$

With this model specification, the default JZS Bayes factor quantifies the extent to which the data support $\mathscr{M}_1$ with path τ' versus $\mathscr{M}_0$ without path τ'.

As before, the resulting JZS Bayes factor for τ' can be converted to a posterior probability:

$$p(\tau' \neq 0 | \mathbf{D}) = \frac{BF_{\tau'}}{BF_{\tau'} + 1}. \tag{23}$$

Together, the Bayes factor for τ' and the Bayes factor for mediation indicate whether mediation is full or partial: If the Bayes factor for mediation is substantially larger than one and the Bayes factor for τ' is substantially smaller than one, there is evidence for full mediation. On the other hand, if the Bayes factor for mediation is substantially larger than one and the Bayes factor for τ' is substantially greater than one, there is evidence for partial mediation.

Simulation study

In order to provide an indication of how the mediation test performs, we designed a simulation study. The goal of the simulation study was to confirm that the Bayes factor draws the correct conclusion: When mediation is present, we expect BF_{med} to be higher than 1; when mediation is absent, we expect BF_{med} to be lower than 1.

Creating the data sets

We assessed performance of the test in different scenarios. The parameters α and β could take the values 0, .30, and .70; τ' was fixed to zero. We did not vary τ' since it has no influence on the Bayes factor for mediation, which only concerns the effect $\alpha\beta$. Furthermore, we chose four sample sizes: N=20, 40, 80, and 160. The 3×3 parameter values combined with the four sample sizes resulted in 36 different scenarios. For each scenario, we created the corresponding covariance matrix of X, Y, and M, all with a variance of one. This standardization has no bearing on the results, since they are scale free. We then used the covariance matrix to generate for each scenario N multivariate normally distributed values for X, M, and Y.[1]

Results

Figure 2 shows the natural logarithm of the Bayes factors for mediation in the different scenarios. The different shades of gray of the panels show the strength of the mediation that governed the generated data: the darker the gray, the stronger the mediation. In the scenarios in which there was no mediation (α=0 and/or β=0), the Bayes factors indicated moderate to very strong evidence for the null model, depending on the sample size. In the scenario of strong mediation (α=.7 and β=.7), the Bayes factors quickly increase from anecdotal evidence (N=20) to moderate evidence (N=40) and further on to very strong and extreme evidence for mediation. In the scenarios of moderate mediation (α=.7 and β=.3 and vice versa), the Bayes factors start to indicate evidence for mediation from sample sizes of around 60. In the scenario of weak mediation (α=.3 and β=.3), the mediation is too weak for the proposed test to detect it with small sample sizes. In those scenarios, the test starts to indicate evidence for mediation only from a sample size of around 80 onward. In summary, the proposed test can distinguish between no mediation and mediation, provided that effect size and sample size are sufficiently large.

Discussion

The results from the simulation study confirm that the JZS Bayesian hypothesis test for mediation performs as advertised: When mediation is absent, the test indicates moderate to strong evidence against mediation, and when mediation is present, the test indicates evidence for mediation, provided that effect size and sample size are sufficiently large. As was expected, the evidence for mediation increases with effect size and with sample size.

Even though the default test performs well in a qualitative sense, it has one shortcoming that remains to be addressed:

[1] We generated data that covaried exactly according to the input covariance matrix. Because the covariances of the data were equal to the covariances of the population, there was no need to control for random sampling, and we simulated only one experiment per scenario. The full simulation code is available in the supplemental materials.

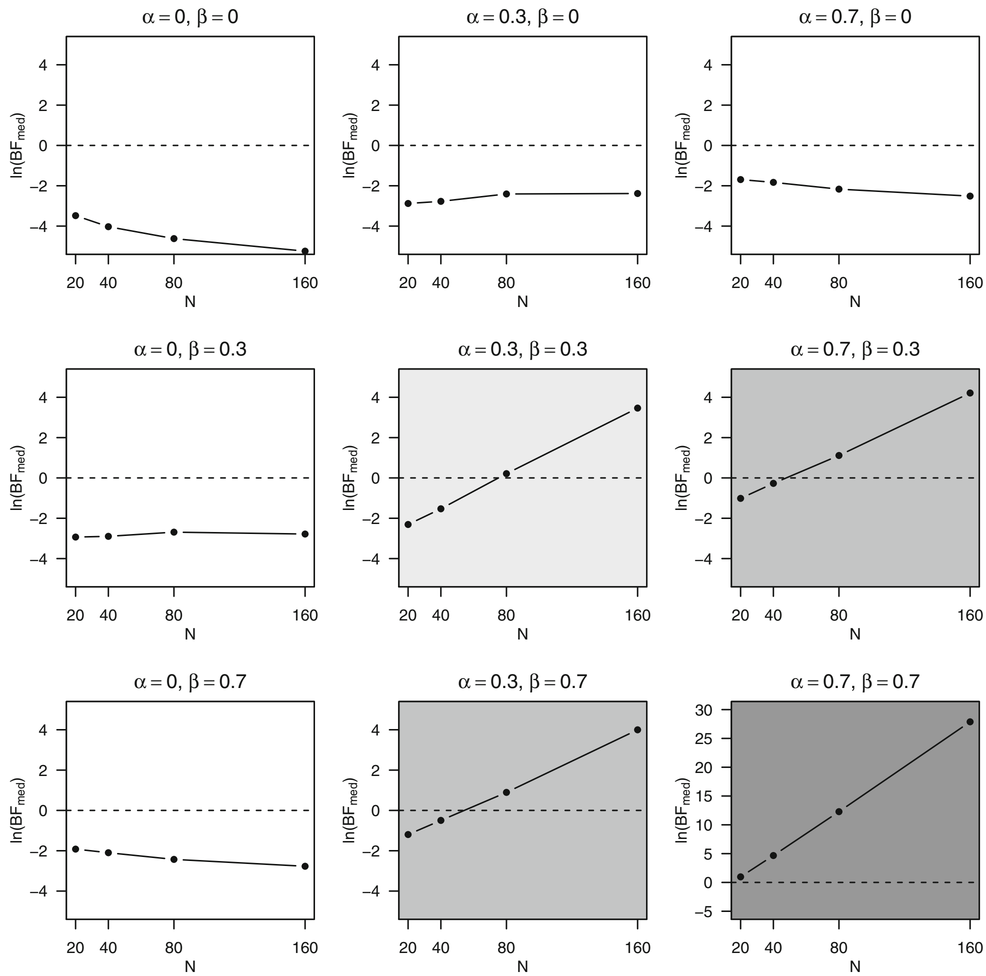


Fig. 2 Performance of the default JZS Bayesian hypothesis test for mediation in different scenarios. Each panel shows the natural logarithm of the Bayes factor for mediation for different values of α and β and different sample sizes. The white panels correspond to scenarios in which there is no mediation; the gray panels to scenarios in which there is mediation. The darker the panel, the stronger the mediation that is present. The horizontal dotted line at zero indicates the boundary that separates evidence for the null model (below the line) and evidence for the mediation model (above the line). Note that the scaling in the scenario of strong mediation is different from the other scenarios to give a more adequate overview of the results

With the proposed method, it is not possible to perform a one-sided test. This is regrettable because, in many situations, the researcher has a clear idea on the direction of the possible paths α, β, and τ'. In order to perform a one-sided Bayesian hypothesis test, the prior needs to be restricted such that it assigns mass to only positive (or negative) values. This is not possible in the mediation test as outlined above.

Extension to one-sided tests

As was mentioned above, our default prior on a regression coefficient is a Cauchy(0,1) distribution. This prior instantiates a two-sided test, since it represents the belief that the effect is just as likely to be positive as negative. In many situations, however, researchers have strong prior ideas about the direction of the effect (Hoijtink et al., 2008). In the Bayesian framework, such prior ideas are directly reflected in the prior distribution. More specifically, suppose that we expect path α to be greater than zero and we seek a test of this order-restricted hypothesis against the null hypothesis that α is zero. For this we consider the following three hypotheses:

$$\mathcal{M}_0 : \alpha = 0,$$
$$\mathcal{M}_1 : \alpha \sim \text{Cauchy}(0, 1),$$
$$\mathcal{M}_2 : \alpha \sim \text{Cauchy}^+(0, 1),$$

where Cauchy$^+$(0,1) indicates that α can take values only on the positive side of the Cauchy(0,1) distribution (i.e., it is a folded Cauchy distribution).

The test of interest features the comparison between the one-sided hypothesis $\mathcal{M}_2$ versus the null hypothesis $\mathcal{M}_0$; that is, we seek the Bayes factor BF_{20}. This Bayes factor can be derived in many ways—for instance, using relatively straightforward techniques such as the Savage-Dickey density ratio (Dickey & Lientz, 1970; Wagenmakers et al., 2010; Wetzels, Grasman, & Wagenmakers, 2010) or relatively intricate techniques such as the reversible jump MCMC (Green, 1995). Here, we apply a different method that is possibly the most reliable and the least computationally expensive (Morey & Wagenmakers, 2014; Pericchi, Liu, & Torres, 2008). This method takes advantage of the fact that we can easily calculate the two-sided Bayes factor, BF_{10}. With this Bayes factor in hand, we only need to apply a simple correction to derive the desired one-sided Bayes factor BF_{10}. Specifically, note that the Bayes factor is transitive:

$$BF_{20} = BF_{21} \times BF_{10}, \tag{24}$$

which is immediately apparent from its expanded form

$$\frac{p(D|\mathscr{M}_2)}{p(D|\mathscr{M}_0)} = \frac{p(D|\mathscr{M}_2)}{p(D|\mathscr{M}_1)} \times \frac{p(D|\mathscr{M}_1)}{p(D|\mathscr{M}_0)}. \tag{25}$$

Thus, the desired one-sided test on α requires only BF_{21} and BF_{10}. We already have access to BF_{10}, and this leaves the calculation of BF_{21}—that is, the Bayes factor in favor of the order-restricted model $\mathscr{M}_2$ over the unrestricted model $\mathscr{M}_1$. As was shown by Klugkist, Laudy, and Hoijtink (2005), this Bayes factor equals the ratio of two probabilities that can be easily obtained: The first is the posterior probability that $\alpha>0$, under the unrestricted model $\mathscr{M}_1$; the second is the prior probability that $\alpha>0$, again under the unrestricted model $\mathscr{M}_1$. Formally,

$$BF_{21} = \frac{p(\alpha > 0|\mathscr{M}_1, \mathbf{D})}{p(\alpha > 0|\mathscr{M}_1)}. \tag{26}$$

Since the prior distribution is symmetric around zero, the denominator equals .5, and Eq. 26 can be further simplified to

$$BF_{21} = 2 \cdot p(\alpha > 0|\mathscr{M}_1, \mathbf{D}) \tag{27}$$

One straightforward way to determine $p(\alpha>0|\mathscr{M}_1,\mathbf{D})$ is (1) to use a generic program for Bayesian inference such as WinBUGS, JAGS, or Stan; (2) implement $\mathscr{M}_1$ in the program and collect Markov chain Monte Carlo (MCMC) samples from the posterior distribution of α; (3) approximate $p(\alpha>0|\mathscr{M}_1,\mathbf{D})$ by the proportion of posterior MCMC samples for α that are greater than zero.[2]

In our implementation of the one-sided mediation tests, we make use of Eqs. 24 and 27. In order to obtain BF_{21}, we implemented the unrestricted models in JAGS (Plummer, 2009). We confirmed the correctness of our JAGS implementation by comparing the analytical results for the two-sided Bayes factor BF_{10} against the Savage-Dickey density ratio results based on the MCMC samples from JAGS (see Appendix 2). The JAGS code itself is provided in Appendix 1, since it allows researchers to adjust the prior distributions if they so desire. Finally, note that our one-sided mediation test can incorporate order-restriction on any of the paths simultaneously.

Example: the firefighter data

To illustrate the workings of the various mediation tests, we will apply them to the same example data Yuan and MacKinnon (2009) used, concerning the PHLAME firefighter study (Elliot et al., 2007). In this study, it was investigated whether the effect of a randomized exposure to one of three interventions (X) on the reported eating of fruits and vegetables (Y) was mediated by knowledge of the benefits of eating fruits and vegetables (M; see Eqs. 1, 2, and 3). The interventions were either a "team-centered peer-led curriculum" or "individual counseling using motivational interviewing techniques," both to promote a healthy lifestyle, or a control condition. The correlation matrix of the data is shown in Table 2.

The conventional approach: the frequentist product method

Yuan and MacKinnon (2009) first reported the results of the conventional frequentist product method mediation analysis. This method tests whether the indirect effect $\alpha\beta$ differs significantly from zero. The estimate for $\alpha\beta$ was .056 with a standard error of .026 (estimated with the Sobel method; Sobel, 1982), with the 95 % confidence interval (.013,.116) (see Table 3) (MacKinnon, Lockwood, & Williams, 2004; the interval takes into account that $\alpha\beta$ is not normally distributed). Since the 95 % confidence interval does not include zero, frequentist custom suggests that the test provides evidence that the effect of X on Y is mediated by M.

The Yuan and MacKinnon (2009) approach: Bayesian parameter estimation

Next, Yuan and MacKinnon (2009) reported the results of their Bayesian mediation analysis, which is based on parameter estimation with noninformative priors. The mean of the posterior distribution of $\alpha\beta$ was .056 with a standard error of .027. The 95 % credible interval for $\alpha\beta$ was (.011,.118) (see Table 3). These Bayesian estimates are numerically consistent with the frequentist results. It should be stressed, however, that the 95 % credible interval does not allow a test. As summarized by Berger (2006), "Bayesians cannot test precise hypotheses using confidence intervals. In classical statistics one frequently sees testing done by forming a confidence region for the parameter, and then rejecting a null value of the parameter if it does not lie in the confidence region. This is simply wrong if done in a Bayesian formulation (and if the null value of the parameter is believable as a hypothesis)" (p. 383; see also Lindley, 1957; Wagenmakers & Grunwald, 2006).

Table 2 Correlation matrix of the PHLAME firefighter data (N=354)

	X	Y	M
X	1.00	.08	.18
Y	.08	1.00	.16
M	.18	.16	1.00

[2] The approximation can be made arbitrarily close by increasing the number of MCMC samples.

Table 3 Three estimates of the mediated effect $\widehat{\alpha}\widehat{\beta}$ for the PHLAME firefighter data set with associated 95 % confidence/credible intervals

	$\widehat{\alpha}\widehat{\beta}$	$CI_{95\%}$
Frequentist product method	.056	(.013, .116)
Yuan & MacKinnon (2009)	.056	(.011, .118)
Default Bayesian hypothesis test	.056	(.012, .116)

The Bayes factor approach: the default Bayesian hypothesis test

We will now consider the results of the proposed Bayesian hypothesis test with the default JZS prior setup. First, we estimated the posterior distribution of $\alpha\beta$, using the method of Yuan and MacKinnon (2009), but now with the JZS prior instead of a noninformative prior. The resulting posterior distribution had a mean of .056 and a 95 % credible interval of (.012,.116) (see Table 3). This is consistent with the results of both the frequentist test and the Bayesian mediation estimation routine of Yuan and Mackinnon. As was expected, the choice of the JZS prior setup over a noninformative prior setup does not much influence the results in terms of parameter estimation.

The advantage of the JZS prior specification is that we can also formally test whether the effect differs from zero. Our analytical test indicates that the Bayes factor for path α is 10.06, which corresponds to a posterior probability of 10.06/(10.06+1)=.91. The Bayes factor for path β is 2.68, which corresponds to a posterior probability of 2.68/(2.68+1)=.73. If we multiply these posterior probabilities, we obtain the posterior probability for mediation: .91×.73=.66.This posterior probability is easily converted to a Bayes factor: .66/(1−.66)=1.94. Hence, the data are about twice as likely under the model with mediation than under the model without mediation. In terms of Jeffreys's evidence categories, this evidence is anecdotal or "not worth more than a bare mention."

It is also possible to include an order-restriction in the mediation model at hand. According to the theory, we expect a positive relation between the mediator "knowledge of the benefits of eating fruits and vegetables" and the dependent variable "the reported eating of fruits and vegetables," or in other words, we expect path β to be greater than zero. If we implement this order-restriction, our test indicates a new Bayes factor for path β of 5.33, with a corresponding posterior probability of 5.33/(5.33+1)=0.84. If we multiply the posterior probability of α with the new posterior probability of β, we obtain the new posterior probability of mediation: .91×.84=.76, with a corresponding Bayes factor for mediation of .76/(1−.76)=3.17. With the imposed order restriction, the observed data are now about three times as likely under the mediation model than under the model without mediation, which according to the Jeffreys' evidence categories constitutes evidence for mediation on the border between "anecdotal" and "moderate."

R package: BayesMed

In order to make our default Bayesian hypothesis tests available, we built the R package BayesMed (Nuijten et al., 2014). R is a free software environment for statistical computing and graphics (R Core Team, 2012), which makes it a good platform for our test.

BayesMed includes both the basic test for mediation (jzs_med) and the accompanying tests for correlation (jzs_cor) and partial correlation (jzs_partcor), as well as the associated Savage-Dickey density ratio versions (jzs_medSD, jzs_corSD, and jzs_partcorSD, respectively). Furthermore, we added the possibility of estimating the indirect effect $\alpha\beta$, based on the procedure outlined in Yuan and MacKinnon (2009), but with a JZS prior setup. Finally, we also included the firefighter data. The use of the tests and their options are described in the help files within the package.

Concluding comments

We have outlined a default Bayesian hypothesis test for mediation and presented an R package that allows it to be applied easily. This default test complements the earlier work by Yuan and MacKinnon (2009) on Bayesian estimation for mediation. In addition, we have extended the default tests by allowing more informative, one-sided alternatives to be tested as well. Nevertheless, our test constitutes only a first step.

A next step could be to extend the test to multiple mediator models. This should be relatively straightforward: The mediation model (Eqs. 2 and 3) needs to be changed to allow multiple mediators. Next, the presence of each path can still be assessed in the same way by calculating the Bayes factor for each path (see Steps 1 and 2 above) and combining the separate Bayes factors into an overall Bayes factor for mediation.

Another extension could be to add a scaling factor to the JZS prior to adjust the spread of the prior distribution.[3] At the moment, the prior includes a Cauchy(0,r=1), but a smaller or larger r would make the prior smaller or wider, respectively.

Other avenues for further development include, but are not limited to, the following: (1) integrate the estimation and testing approaches by using the estimation outcomes from earlier work as a prior for the later test (Verhagen & Wagenmakers, in press); (2) explore methods to incorporate substantive prior knowledge (e.g., Dienes, 2011); (3) extend the test to interval null hypotheses—that is, null hypotheses that are not defined by a point mass at zero but, instead, by a practically meaningful area around zero (Morey & Rouder, 2011); and (4) generalize the methodology to more complex models such as hierarchical models or mixture models.

[3] We thank an anonymous reviewer for pointing this out to us.

As for all Bayesian hypothesis tests that are based on Bayes factors, users need to realize that the test depends on the specification of the alternative hypothesis. In general, it is a good idea to conduct a sensitivity analysis and examine the extent to which the outcomes are qualitatively robust to alternative plausible prior specifications (e.g., Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011). Such sensitivity analyses are facilitated by our JAGS code presented in Appendix 1.

In sum, we have provided a default Bayesian hypothesis test for mediation. This test allows users to quantify statistical evidence in favor of both the null hypothesis (i.e., no mediation) and the alternative hypothesis (i.e., full or partial mediation). The test also allows informative hypotheses to be tested in the form of order-restrictions. Several extensions of the methodology are possible and await future implementation.

Acknowledgements This research was supported by an ERC grant from the European Research Council. Conor V. Dolan is supported by the European Research Council (Genetics of Mental Illness; grant number: ERC–230374). Ruud Wetzels is supported by the Dutch national program COMMIT.

Appendixes

Appendix 1. JAGS code

JAGS code for correlation

```
####### Cauchy-prior on alpha #######
model
{
 for (i in 1:n)
  {
    mu[i] <- intercept + alpha*x[i]
    y[i]   ~ dnorm(mu[i],phi)
  }

# uninformative prior on intercept,
# Jeffreys' prior on precision phi
  intercept ~ dnorm(0,.0001)
  phi   ~ dgamma(.0001,.0001)


# inverse-gamma prior on g:
  g       <- 1/invg
  a.gamma <- 1/2
  b.gamma <- n/2
  invg     ~ dgamma(a.gamma,b.gamma)


# g-prior on beta:
  vari <- (g/phi) * invSigma
  prec <- 1/vari
  alpha    ~ dnorm(0, prec)
}

# Explanation---------------------------------
    # Prior on g:
    # We know that g ~ inverse_gamma(1/2, n/2), with 1/2 the shape
    # parameter and n/2 the scale parameter.
    # It follows that 1/g ~ gamma(1/2, 2/n).
    # However, BUGS/JAGS uses the *rate parameterization*
    # 1/theta instead of the scale parametrization theta.
    # Hence we obtain, in de BUGS/JAGS rate notation:
    # 1/g ~ dgamma(1/2, n/2)
    #---------------------------------------
```

JAGS code for partial correlation

```
####### Cauchy-prior on beta and tau' #######

# theta contains beta and tau'

model
{
  for (i in 1:n)
   {
     mu[i] <- intercept + theta[1]*x[i,1] + theta[2]*x[i,2]
     y[i]   ~ dnorm(mu[i],phi)
   }

# uninformative prior on intercept,
# Jeffreys' prior on precision phi
  intercept ~ dnorm(0,.0001)
  phi   ~ dgamma(.0001,.0001)


# inverse-gamma prior on g:
  g       <- 1/invg
  a.gamma <- 1/2
  b.gamma <- n/2
  invg     ~ dgamma(a.gamma,b.gamma)


# calculation of the inverse matrix of V
  inverse.V <- inverse(V)
# calculation of the elements of prior precision matrix
  for(i in 1:2)
  {
    for (j in 1:2)
    {
      prior.T[i,j] <- inverse.V[i,j] * phi/g
    }
  }
# multivariate prior for the theta vector
  theta[1:2] ~ dmnorm( mu.theta, prior.T )
  for(i in 1:2) { mu.theta[i] <- 0 }

}
```

Appendix 2. Testing the correctness of our JAGS implementation

To assess the correctness of our JAGS implementation, we compared the analytical results for the two-sided Bayes factor against the Savage-Dickey density ratio results based on the MCMC samples from JAGS. The distribution that fit the posterior samples best[4] is the nonstandardized *t*-distribution with the following density:

$$p(x|\nu,\mu,\sigma) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\pi\nu\sigma)}}\left(1+\frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}, \qquad (28)$$

with ν degrees of freedom, location parameter μ, and scale parameter σ. With the samples of the parameter of interest, we can estimate ν, μ, and σ and, thus, the exact shape of the distribution and the exact height of the distribution at the point of interest.

We checked the fit of this distribution and the performance of the SD method in a small simulation study. We considered the following sample sizes: N = 20, 40, 80, or 160. We simulated correlational data by drawing N values for X from a standard normal distribution, and conditional on X, we simulated values for Y according to the following equation:

[4] We compared the fit of four distributions: a nonstandardized *t*-distribution, a normal distribution, a nonparametric distribution estimated with the spline interpolation function splinefun in R, and a nonparametric distribution estimated with the R function logspline that also uses splines to estimate the log density. All four distributions fitted reasonably well: The Bayes factors of the analytical test and the SD method are similar with all different posterior distributions. All four distributions are therefore included in the R package BayesMed and can be used when applying the SD method.

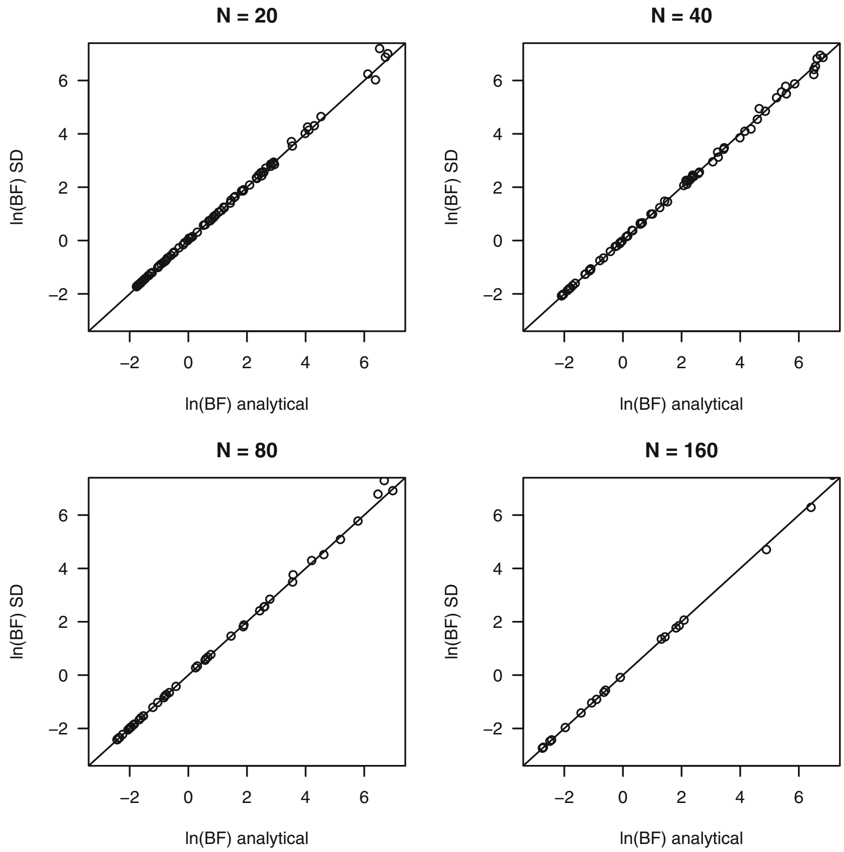


Fig. 3 Natural logarithm of the Bayes factors for correlation obtained with analytical calculations (*x* axis) or obtained with the SD method based on a nonstandardized *t*-distribution (*y* axis) for different sample sizes (*N*). The graphs show fewer points as the samples grow larger, because in these situations, there are more extreme Bayes factors that fall outside the axis limits. We restricted the graphs, since it is most important that the lower Bayes factors lie on the diagonal; it is not important whether a Bayes factor is 2,000 or 3,000, since it is overwhelming evidence in any case

$$Y_i = \beta_0 + \tau X_i + \varepsilon, \quad (29)$$

where the subscript i denotes subject i and τ represents the relation between X and Y. For each of the four sample sizes, we generated 100 data sets, in each of which τ was drawn from a standard uniform distribution.

Next, we tested the correlation in each data set with both the analytical Bayesian correlation test and the SD method with the nonstandardized *t*-distribution and compared the results. The results are shown in Fig. 3. The figure shows that the proposed SD method performs well: The Bayes factors of the analytical test and the SD method are similar for all sample sizes and correlations.

References

Armstrong, A. M., & Dienes, Z. (2013). Subliminal understanding of negation: Unconscious control by subliminal processing of word pairs. *Consciousness and Cognition, 22,* 1022–1040.

Berger, J. O. (2006). Bayes factors. In S. Kotz, N. Balakrishnan, C. Read, B. Vidakovic, & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences, vol. 1* (2nd ed., pp. 378–386). Hoboken, NJ: Wiley.

Berger, J. O., & Delampady, M. (1987). Testing precise hypotheses. *Statistical Science, 2,* 317–352.

Berger, J. O., & Wolpert, R. L. (1988). *The likelihood principle* (2nd ed.). Hayward (CA): Institute of Mathematical Statistics.

Consonni, G., Forster, J. J., & La Rocca, L. (2013). The whetstone and the alum block: Balanced objective Bayesian comparison of nested models for discrete data. *Statistical Science, 28,* 398–423.

Dickey, J. M., & Lientz, B. P. (1970). The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain. *Annals of Mathematical Statistics, 41,* 214–226.

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. New York: Palgrave MacMillan.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? Perspectives on psychological. *Science, 6,* 274–290.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70,* 193–242.

Elliot, D. L., Goldberg, L., Kuehl, K. S., Moe, E. L., Breger, R. K., & Pickering, M. A. (2007). The phlame (promoting healthy lifestyles: Alternative models' effects) firefighter study: Outcomes of two models of behavior change. *Journal of Occupational and Environmental Medicine, 49*(2), 204–213.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika, 82,* 711–732.

Guo, X., Li, F., Yang, Z., & Dienes, Z. (2013). Bidirectional transfer between metaphorical related domains in implicit learning of form-meaning connections. *PLoS ONE, 8,* e68100.

Hoijtink, H., Klugkist, I., & Boelen, P. (2008). *Bayesian evaluation of informative hypotheses*. New York: Springer.

Iverson, G. J., Wagenmakers, E. J., & Lee, M. D. (2010). A model averaging approach to replication: The case of p_{rep}. *Psychological Methods, 15,* 172–181.

Jeffreys, H. (1961). *Theory of Probability* (3rd ed.). Oxford, UK: Oxford University Press

Kass, R. E., & Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association, 90,* 928–934.

Klugkist, I., Laudy, O., & Hoijtink, H. (2005). Inequality constrained analysis of variance: A Bayesian approach. *Psychological Methods, 10,* 477.

Kruschke, J. K. (2010). *Doing Bayesian data analysis: A tutorial introduction with R and BUGS*. Burlington, MA: Academic Press.

Lee, M. D., & Wagenmakers, E. J. (2013). *Bayesian modeling for cognitive science: A practical course*. Germany: Cambridge University Press.

Lewis, S. M., & Raftery, A. E. (1997). Estimating Bayes factors via posterior simulation with the Laplace–Metropolis estimator. *Journal of the American Statistical Association, 92,* 648–655.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., & Berger, J. O. (2008). Mixtures of *g* priors for Bayesian variable selection. *Journal of the American Statistical Association, 103,* 410–423.

Lindley, D. V. (1957). A statistical paradox. *Biometrika, 44,* 187–192.

MacKinnon, D. P., Fairchild, A., & Fritz, M. (2007). Mediation analysis. *Annual Review of Psychology, 58,* 593.

MacKinnon, D. P., Lockwood, C. M., & Hoffman, J. (1998). *A new method to test for mediation*. Paper presented at the annual meeting of the Society for Prevention Research, Park City, UT.

MacKinnon, D. P., Lockwood, C., Hoffman, J., West, S., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7,* 83–104.

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

MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research, 30,* 41–62.

Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods, 16,* 406–419.

Morey, R. D., & Wagenmakers, E. J. (2014). *Simple relation between one–sided and two–sided Bayesian point–null hypothesis tests*. Manuscript submitted for publication.

Myung, I. J., & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review, 4,* 79–95.

Nuijten, M. B., Wetzels, R., Matzke, D., Dolan, C. V., & Wagenmakers, E. J. (2014). *BayesMed: Default Bayesian hypothesis tests for correlation, partial correlation, and mediation*. R package version 1.0. http://CRAN.R-project.org/package=BayesMed

O'Hagan, A., & Forster, J. (2004). *Kendall's advanced theory of statistics vol. 2B: Bayesian inference* (2nd ed.). London: Arnold.

Overstall, A. M., & Forster, J. J. (2010). Default Bayesian model determination methods for generalised linear mixed models. *Computational Statistics & Data Analysis, 54,* 3269–3288.

Pericchi, L. R., Liu, G., & Torres, D. (2008). Objective Bayes factors for informative hypotheses: "Completing" the informative hypothesis and "splitting" the Bayes factor. In H. Hoijtink, I. Klugkist, & P. A. Boelen (Eds.), *Bayesian evaluation of informative hypotheses* (pp. 131–154). New York: Springer Verlag.

Plummer, M. (2009). *JAGS version 1.0. 3 manual*. URL: http://www-ice.iarc.fr/~martyn/software/jags/jags_user_manual. pdf

R Core Team. (2012). *R: A language and environment for statistical computing* []. Vienna, Austria. APACrefURL http://www.R-project.org/ ISBN 3-900051-07-0

Rouder, J. N., & Morey, R. D. (2012). Default Bayes factors for model selection in regression. *Multivariate Behavioral Research, 47,* 877–903.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56,* 356–374.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16,* 225–237.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461–464.

Sellke, T., Bayarri, M. J., & Berger, J. O. (2001). Calibration of *p* values for testing precise null hypotheses. *The American Statistician, 55,* 62–71.

Semmens-Wheeler, R., Dienes, Z., & Duka, T. (2013). Alcohol increases hypnotic susceptibility. *Consciousness and Cognition, 22*(3), 1082–1091.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology, 13,* 290–312.

Vandekerckhove, J, Matzke, D., & Wagenmakers, E. J. (in press). Model comparison and the principle of parsimony. In J. Busemeyer, J. Townsend, Z. J. Wang, & A. Eidels (Eds.), *Oxford handbook of computational and mathematical psychology*. Oxford University Press.

Venzon, D., & Moolgavkar, S. (1988). A method for computing profile-likelihood-based confidence intervals. *Applied Statistics, 37*(1), 87–94.

Verhagen, J., & Wagenmakers, E. J. (in press). A Bayesian test to quantify the success or failure of a replication attempt. *Journal of Experimental Psychology: General*.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review, 14,* 779–804.

Wagenmakers, E. J., & Grünwald, P. (2006). A Bayesian perspective on hypothesis testing. *Psychological Science, 17,* 641–642.

Wagenmakers, E. J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology, 60,* 158–189.

Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi. *Journal of Personality and Social Psychology, 100,* 426–432.

Wetzels, R., Grasman, R. P. P. P., & Wagenmakers, E. J. (2010). An encompassing prior generalization of the Savage–Dickey density ratio test. *Computational Statistics & Data Analysis, 54,* 2094–2102.

Wetzels, R., Grasman, R. P. P. P., & Wagenmakers, E. J. (2012). A default Bayesian hypothesis test for ANOVA designs. *The American Statistician, 66,* 104–111.

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 *t* tests. *Perspectives on Psychological Science, 6,* 291–298.

Wetzels, R., Raaijmakers, J. G. W., Jakab, E., & Wagenmakers, E. J. (2009). How to quantify support for and against the null hypothesis: A flexible WinBUGS implementation of a default Bayesian *t* test. *Psychonomic Bulletin & Review, 16,* 752–760.

Wetzels, R., & Wagenmakers, E. J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review, 19,* 1057–1064.

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

Zellner, A., & Siow, A. (1980). Posterior odds ratios for selected regression hypotheses. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, & A. F. M. Smith (Eds), *Bayesian statistics* (pp. 585–603). Valencia: University Press.