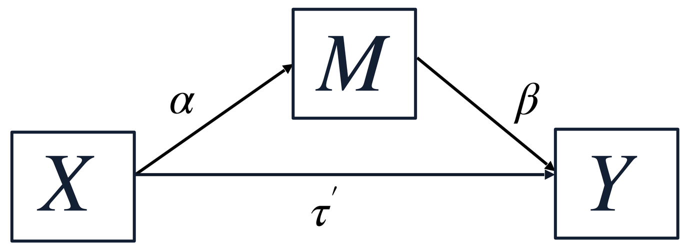
# Evaluating the Performance of the Bayes Factor for Testing Mediation Effects

**Xiao Liu[1] (xiao.liu@austin.utexas.edu), Zhiyong Zhang[2], and Lijuan Wang[2]**

[1]Department of Educational Psychology, The University of Texas at Austin [2]Department of Psychology, University of Notre Dame

## Introduction

Frequently of interest: ***Testing*** the presence vs. absence of a ***mediation*** effect of a treatment $X$ (e.g., intervention assignment, antecedent variable) on an outcome $Y$ via a hypothesized mediator $M$.




**Most of the tests** for mediation: NHST (null hypothesis significance testing)

- e.g.: Sobel test, joint significance test, interval-based tests (e.g., bootstrap interval, credible interval).

***Bayes factor (BF)*** — Promising alternative to NHST

- Increasingly popular in psychology (Heck et al., 2022).
- Relative likelihood of an alternative hypothesis ($H_1$) to a null hypothesis ($H_0$)

$$\mathrm{BF} = \frac{p(\mathrm{data}\,|\,H_1)}{p(\mathrm{data}\,|\,H_0)}$$

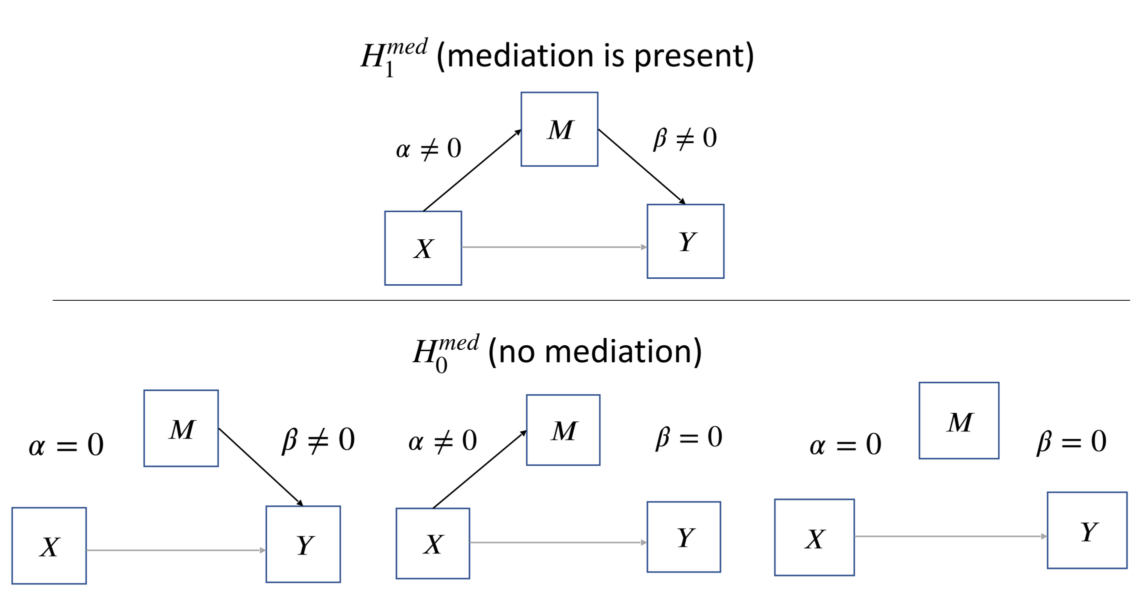
- Compared to NHST, the BF has several advantages for hypothesis testing (e.g., Dienes & Mclatchie, 2018), e.g.:
  - The support in data for the null $H_0$ — evaluated.

### The BF for mediation

- Little development. A recently proposed one (Liu et al., 2022):

$$\mathrm{BF}^{med} = \frac{p(\mathrm{data}\,|\,H_1^{med})}{p(\mathrm{data}\,|\,H_0^{med})} = \frac{p(\mathrm{data}\,|\,\alpha \neq 0, \beta \neq 0)}{p(\mathrm{data}\,|\,\alpha = 0, \beta \neq 0)p_{01|0} + p(\mathrm{data}\,|\,\alpha \neq 0, \beta = 0)p_{10|0} + p(\mathrm{data}\,|\,\alpha = 0, \beta = 0)p_{00|0}}$$

- $p_{01|0}, p_{10|0}, p_{00|0}$ are the conditional prior probabilities of the three no-mediation scenarios under the null of no mediation ($H_0^{med}$); specified based on prior knowledge (e.g., all equal 1/3: the three no-mediation scenarios are equally likely to occur based on prior knowledge)




For the **simple mediation model** where the paths $\alpha$ and $\beta$ are independent:

$$\mathrm{BF}^{med} = \frac{(1 + \mathrm{PriorOdds}^{\beta} + \mathrm{PriorOdds}^{\alpha})\mathrm{BF}^{\alpha}\mathrm{BF}^{\beta}}{1 + \mathrm{PriorOdds}^{\beta}\mathrm{BF}^{\beta} + \mathrm{PriorOdds}^{\alpha}\mathrm{BF}^{\alpha}} \quad (1)$$

- $\mathrm{PriorOdds}^{\alpha}$: Prior odds of the presence of path $\alpha$
- $\mathrm{PriorOdds}^{\beta}$: Prior odds of the presence of path $\beta$
- $\mathrm{BF}^{\alpha}, \mathrm{BF}^{\beta}$: BFs for the regression coefficients representing the paths $\alpha$ and $\beta$; can get from many packages, e.g., R package "BayesFactor" (Morey et al., 2018).

## Aims

The BF approach provides an appealing complementary method for testing mediation. Particularly, with $\mathrm{BF}^{med}$:

- The support in data for the specified no-mediation hypothesis $H_0^{med}$ is evaluated.
- Prior probabilities of the three no-mediation scenarios can be considered.

***But***: How well does $\mathrm{BF}^{med}$ perform for testing mediation, in terms of

- *true positive rate* & *false positive rate* (i.e., power & Type I error rate in the frequentist language)?
- Particularly: Impacts of the prior specification? (e.g., can careful use of prior knowledge benefit the performance?)

We focus on the simple mediation model.

***Challenge***: Method for calculation of the true- or false-positive rates, which are

True positive rate $= \Pr(\mathrm{BF}^{med} > \mathrm{cutoff}\,|\,H_1^{med})$

False positive rate $= \Pr(\mathrm{BF}^{med} > \mathrm{cutoff}\,|\,H_0^{med})$

- Need the distributions of $\mathrm{BF}^{med}$ over repeated samples under $H_0^{med}$ and $H_1^{med}$.
- Need the cutoff.

## Proposed Simulation-Based Method

Two types of priors can be distinguished when obtaining the distributions of a BF (Schönbrodt & Wagenmakers, 2018):

- Analysis prior: used to calculate BFs;
- Design prior: used to specify the population under a hypothesis (e.g., point mass at 0 under the null of no effect; a fixed effect size under an alternative)

For the BF for testing mediation, we further have

- *Analysis prior odds* of each path ($\alpha$ or $\beta$) used to calculate the $\mathrm{BF}^{med}$ with Eq.(1)
- *Design prior odds* of each path ($\alpha$ or $\beta$) used to specify the population under no-mediation ($H_0^{med}$),
  - i.e., used to determine the probability of the population being in each of the three no-mediation scenarios

**Simulating the distribution of $\mathrm{BF}^{med}$ under $H_0^{med}$**

Step.1: Specify the population under $H_0^{med}$, including population parameter values ($\alpha, \beta, \tau'$) and the design prior odds.

Step.2: Simulate a sample of size $n$ from the population under $H_0^{med}$

Step.3: Calculate the $\mathrm{BF}^{med}$ with Eq.(1) using the analysis prior (e.g., default prior in "BayesFactor" package, "default" prior odds: $\mathrm{PriorOdds}^{\alpha} = \mathrm{PriorOdds}^{\beta} = 1$)

Step.4: Repeat Steps 1-3 many (e.g., 10000) times.

**Simulating the distribution of $\mathrm{BF}^{med}$ under $H_1^{med}$**

Step.1': Specify the population under $H_1^{med}$, including population parameter values ($\alpha, \beta, \tau'$), where $\alpha, \beta$ are non-zero.

Step.2': Simulate a sample of size $n$ from the population under $H_1^{med}$

Step.3': same as Step.3; Step.4': Repeat Steps 1'-3' many (e.g., 10000) times.

**Cutoff of the BF for testing mediation: Two methods**

- *Absolute cutoff*: 3 is a common cutoff used in previous studies on BFs of single relations (e.g., a regression coefficient; e.g., Jeon & De Boeck, 2017).
- *Relative cutoff* for 5% false positive ("Type I error") rate: 95% quantile of the distribution of $\mathrm{BF}^{med}$ under $H_0^{med}$.

## True (false) positive rates of the BF for mediation

The BF with two different methods of specifying the analysis prior odds:

(1) $\mathrm{BF}^{med}$ [design]: analysis prior odds = design prior odds;

(2) $\mathrm{BF}^{med}$ [default]: analysis prior odds = the default prior odds (1 for both paths $\alpha$ and $\beta$, so $\mathrm{BF}^{med}\,[\mathrm{default}] = \frac{3\mathrm{BF}^{\alpha}\mathrm{BF}^{\beta}}{1 + \mathrm{BF}^{\alpha} + \mathrm{BF}^{\beta}}$).

Also, a frequentist mediation test (the joint-significance test) for comparison, for which:

- the power (and Type I error rate) = the proportions of times that the absence of mediation was rejected with the samples generated under $H_1^{med}$ (and under $H_0^{med}$).

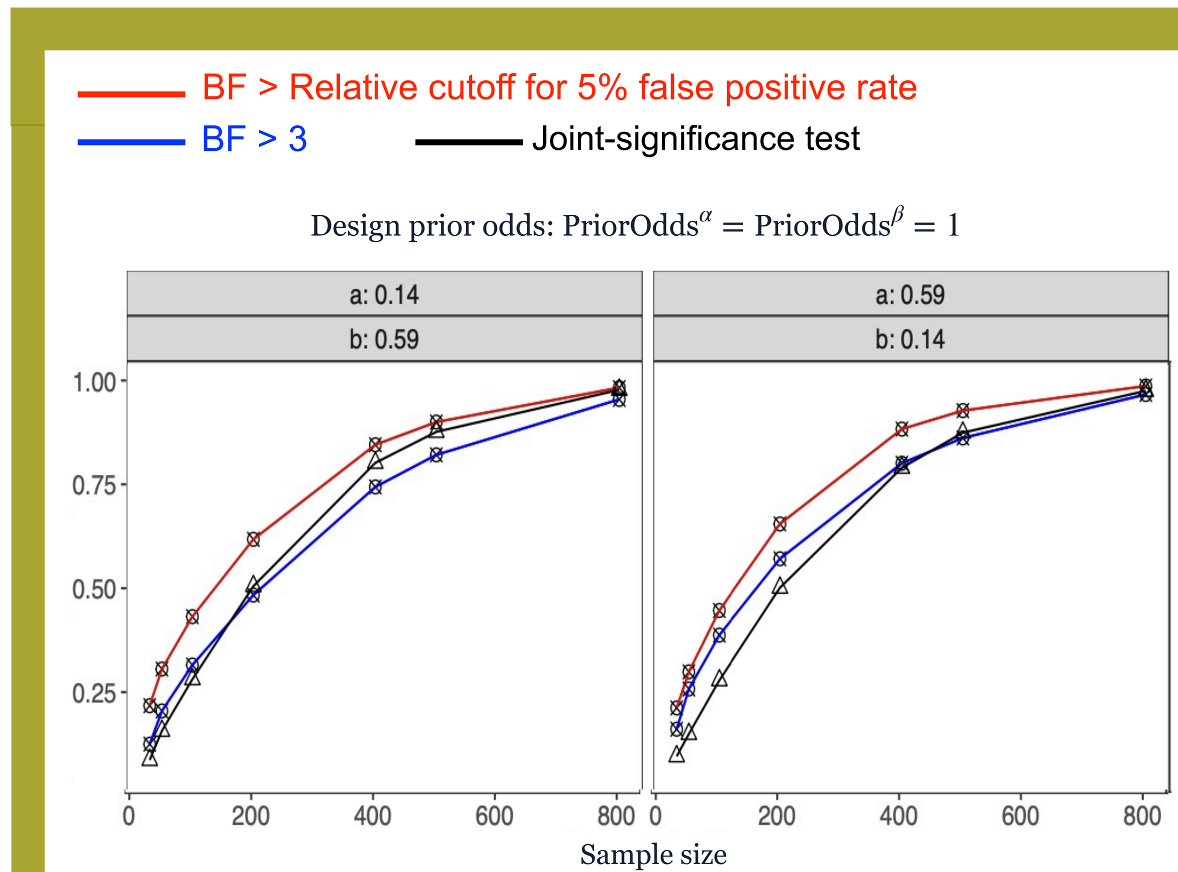
Sample size: $n = 50$

| Design prior odds | BF cutoff: Relative | | | | BF cutoff: 3 | | | | Joint-significance test | |
|---|---|---|---|---|---|---|---|---|---|---|
| | True positive rates $\mathrm{BF}^{med}$ | | Cutoffs for 5% false positive rates $\mathrm{BF}^{med}$ | | True positive rates $\mathrm{BF}^{med}$ | | False positive rates $\mathrm{BF}^{med}$ | | Power | Type I error rate |
| $\mathrm{PriorOdds}^{\alpha} = 1$ | [design] | [default] | [design] | [default] | [design] | [default] | [design] | [default] | | |
| $\mathrm{PriorOdds}^{\beta}$ | True effect size: $\alpha = 0.59$, $\beta = 0.14$ | | | | | | | | | |
| 0.01 | 0.23 | 0.25 | 2.43 | 3.13 | 0.18 | 0.26 | 0.04 | 0.05 | 0.15 | 0.03 |
| 0.33 | 0.26 | 0.26 | 2.43 | 2.97 | 0.21 | | 0.04 | 0.05 | | 0.02 |
| 1 | 0.30 | 0.30 | 2.61 | 2.61 | 0.26 | | 0.04 | 0.04 | | 0.02 |
| 3 | 0.41 | 0.39 | 3.09 | 2.07 | 0.42 | | 0.05 | 0.03 | | 0.02 |
| 100 | **0.97** | **0.67** | **1.68** | **1.31** | **0.93** | | **0.03** | **0.01** | | **0.01** |
| | True effect size: $\alpha = 0.14$, $\beta = 0.59$ | | | | | | | | | |
| 0.01 | **0.96** | **0.54** | 1.51 | **1.16** | **0.92** | 0.20 | **0.02** | **0.00** | 0.16 | **0.01** |
| 0.33 | 0.49 | 0.41 | 2.72 | 1.51 | 0.44 | | 0.04 | 0.02 | | **0.01** |
| 1 | 0.31 | 0.31 | 2.01 | 2.01 | 0.20 | | 0.03 | 0.03 | | 0.02 |
| 3 | 0.20 | 0.22 | 1.86 | 2.86 | 0.12 | | **0.02** | 0.05 | | 0.03 |
| 100 | 0.17 | 0.17 | **1.41** | 3.71 | 0.09 | | **0.02** | 0.07 | | 0.05 |

***Key findings***:

1. The prior odds specifications impacted the performance (true/false positive rates) of the BF for mediation, depending on the relative true effect sizes of $\alpha$ vs. $\beta$.
   - E.g., when path $\beta$ had a relatively small true effect size (i.e., $\beta < \alpha$): the true positive rates of $\mathrm{BF}^{med}$ [design] was
     - increased with specifying larger prior odds for path $\beta$;
     - higher than the true positive rate of $\mathrm{BF}^{med}$ [default] and the power of the joint-significance test, with specifying a sufficiently large prior odds for path $\beta$.
   - Minor impact of the prior odds specification on the true positive rates, when the two paths had similar true effect sizes (i.e., $\beta = \alpha$; results not shown here).

2. The cutoff defining methods impacted the true positive rate of the BF for mediation.




Generally, for both $\mathrm{BF}^{med}$ [default] and [design],

- the relative cutoffs for 5% false positive rates <3; thus,
- higher true positive rates with the relative cutoffs, compared to with the cutoff 3.
- Consistent with previous research on the BFs for regression coefficients (e.g., Jeon & De Boeck, 2017; Rouder et al., 2009).

## REFERENCES

Dienes, Z., & Mclatchie, N. (2018). Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic Bulletin & Review*, 25(1), 207–218.
Heck, D. W., Boehm, U., Böing-Messing, F., Bürkner, P.-C., Derks, K., Dienes, Z., . . . others (2022). A review of applications of the bayes factor in psychological research. *Psychological Methods,* 28(3), 558–579.
Jeffreys, H. (1961). Theory of probability (3rd ed.). UK: Oxford University Press: Oxford.
Liu, X., Zhang, Z., & Wang, L. (2023). Bayesian hypothesis testing of mediation: Methods and the impact of prior odds specifications. *Behavior Research Methods*, 55(3), 1108–1120.
Morey, R. D., Rouder, J. N., & Jamil, T. (2018). BayesFactor: Computation of Bayes Factors for common designs. R package version 0.9. 12-4.2.
Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, 25(1), 128–142.
Jeon, M., & De Boeck, P. (2017). Decision qualities of bayes factor and p value-based hypothesis testing. *Psychological Methods*, 22(2), 340-360.
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(2), 225–237.