Questions of value, questions of magnitude: An exploration and application of methods for comparing indirect effects in multiple mediator models

Jacob J. Coutts[1] · Andrew F. Hayes[2]

Abstract

Mediation analysis is widely used to test and inform theory and debate about the mechanism(s) by which causal effects operate, quantitatively operationalized as an *indirect effect* in a mediation model. Most effects operate through multiple mechanisms simultaneously, and a mediation model is likely to be more realistic when it is specified to capture multiple mechanisms at the same time with the inclusion of more than one mediator in the model. This also allows an investigator to compare indirect effects to each other. After an overview of the mechanics of mediation analysis, we advocate formally comparing indirect effects in models that include more than one mediator, focusing on the important distinction between questions and claims about *value* (i.e., are two indirect effects the same number?) versus *magnitude* (i.e., are two indirect effects equidistant from zero or the same in strength?). After discussing the shortcomings of the conventional method for comparing two indirect effects in a multiple mediator model—which only answers a question about magnitude in some circumstances—we introduce several methods that, unlike the conventional approach, always answer questions about difference in magnitude. We illustrate the use of these methods and provide code that implements them in popular software. We end by summarizing simulation findings and recommending which method(s) to prefer when comparing like- and opposite-signed indirect effects.

Keywords Comparing indirect effects · Multiple mediator models · Contrasts · Indirect effects · Simulation study

Behavioral scientists have long sought to explore and understand causal relationships between variables. Knowing that one variable causes another provides valuable theoretical and practical information to the scientific community and users of its findings. For instance, if a self-compassion training program causes well-being to improve, it can become central to behavioral or therapeutic interventions that aim to improve an individual's mental health. But knowing that one variable causally influences another is only part of the story scientists seek to tell. Also important to the narrative is *how* that effect operates. For example, perhaps self-compassion increases well-being because greater self-compassion first causes greater optimism, which, in turn, causes well-being to improve. That is, perhaps self-compassion is acting *indirectly, through* optimism, to enhance one's well-being.

Mediation analysis is frequently used to understand how some presumed causal antecedent variable (hereinafter, *X*) influences an outcome of interest (hereinafter, *Y*) by affecting one or more *mediator*(s) (hereinafter, *M*) that, in turn, influence the outcome. Mediation models—which focus primarily on the estimation of indirect effects of *X* on *Y*—are abundant in behavioral science research. For instance, in one study, Schönfeld et al. (2016) found that daily stress had a significant and negative effect on a person's mental health. The researchers posited that more stress would reduce a person's level of self-efficacy, and this reduction in self-efficacy is what causes poorer mental health. In other words, more stress would indirectly influence mental health through self-efficacy. A mediation analysis produced results consistent with this explanation. This is merely one of many examples of mediation analysis conducted in research. Mediation models have been applied in a number of fields, including communication (Hoffman & Young, 2011), public health (Ho et al., 2013), education (Jin et al., 2018), nursing (Van der Heijden et al., 2019), and business (Wieder & Ossimitz, 2015), among many other disciplines.

✉ Jacob J. Coutts
coutts.14@osu.edu

1 The Ohio State University, Columbus, OH, USA

2 University of Calgary, Calgary, AB, Canada

But real-world causal processes are typically complex, and it is unlikely that just one mechanism is transmitting X's effect on Y. Rather, it is more plausible that X has effects on several mediating variables causally downstream, each of which then influences Y. For instance, self-compassion training (X) could cause both greater optimism (M_1) and reduced rumination (M_2), both of which then influence mental health (Y). Simultaneously estimating multiple mechanisms by which X affects Y (i.e., multiple indirect effects) can be beneficial to researchers for a variety of reasons, chief among them being that researchers can test competing theories against each other by comparing indirect effects through different mediators. For example, if both optimism and rumination were measured in the same study and included in a single mediation model, it would be possible to test if there is a difference between the effects of self-compassion training on mental health through optimism and through rumination to see whether one is larger, dominant, or "more important" in some sense in explaining the effect of X on Y.

A few techniques for comparing indirect effects have been discussed in the mediation analysis literature (e.g., MacKinnon, 2000; Preacher & Hayes, 2008; Raykov et al., 2008). However, the methods most widely used suffer from the limitation that they aren't always appropriate when a researcher's question focuses on whether the indirect effects are equal in *magnitude* rather than *value*. After reviewing the mechanics of the estimation of indirect effects in a mediation analysis, this paper contributes to the literature by first making the case for the need to compare indirect effects operating through different mediators. We then talk about the most common (albeit, regrettably rare) practice for comparing indirect effects, highlighting its shortcomings in situations in which two mechanisms are working in competition to each other. We then describe several alternative methods for comparing indirect effects that can be used to answer questions about magnitude and talk about the results of a simulation study examining their relative performance in situations where the indirect effects are consistent or opposing in direction. We also provide a concrete substantive example of the application of these methods while illustrating their implementation in statistical software.

Indirect effects in mediation models

Mediation models are typically estimated in one of two ways: structural equation modeling (SEM) or through a series of ordinary least squares (OLS) regression analyses. Although some researchers have taken a stance about which approach should be used to estimate mediation models (Iacobucci et al., 2007; Pek & Hoyle, 2016), observed-variable SEM and OLS regression produce largely identical results (Hayes et al., 2017; Rijnhart et al., 2017). There certainly are pros and cons to
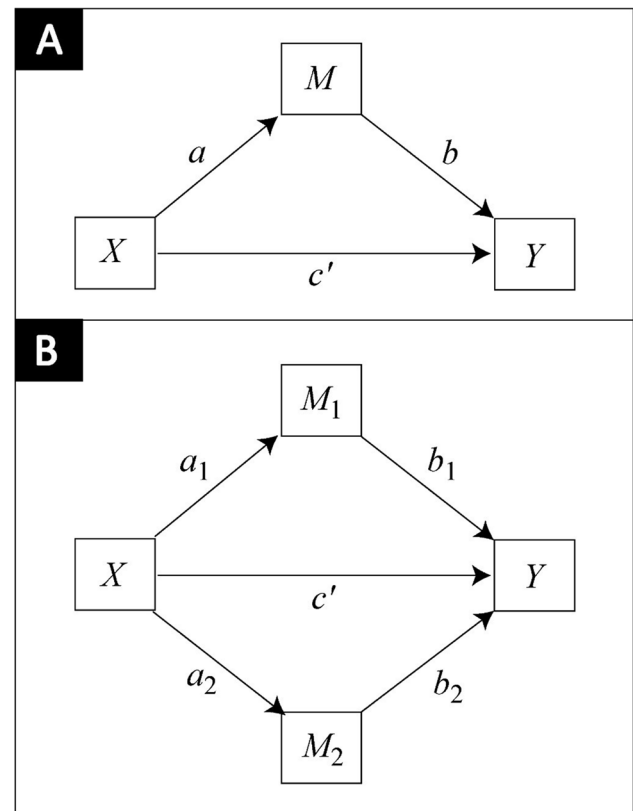
Fig. 1 A simple (**A**) and a parallel multiple mediator model (**B**)

conducting mediation analysis through simultaneous estimation using maximum likelihood in SEM as opposed to a set of OLS regression analyses, but which is used is inconsequential to the difference(s) between indirect effects and the methods we discuss here can be used in either computational arena.

In a mediation model of X's effect on Y through a single mediator M, assuming M and Y are continuous variables and X is either dichotomous or continuous, the effect of X operating *directly* and *indirectly* through M can be estimated using two linear equations, one for M and one for Y:

$$M = d_M + aX + \varepsilon_M \tag{1}$$

$$Y = d_Y + c'X + bM + \varepsilon_Y \tag{2}$$

Figure 1A shows a statistical diagram of this model. In Eq. (1), a represents the estimated difference in M between two observations that differ by one unit on X. In Eq. (2), b represents the difference in Y between two observations that differ by one unit on M but who have the same X value and c' represents the difference in Y between two observations that differ by one unit on X but have the same M value. Assuming proper model specification and the assumptions of valid causal inference are met, these regression weights are typically interpreted as representing causal effects (of X on M, M on Y controlling for X,

and X and Y controlling for M, respectively). The *indirect effect* of X on Y, the effect of most interest in mediation analysis (the effect of X on Y through M), is obtained by multiplying a and b together (i.e., ab). The product ab represents the estimated difference in Y attributable to a one-unit difference in X that operates through the joint causal effect of X on M which in turn affects Y. The weight for X in Eq. (2), c', is the direct effect of X on Y, meaning the effect of X on Y that does not operate through the mechanism represented by the mediator, M.

Statistical evidence consistent with mediation is found in an inference that the indirect effect is different from zero by some kind of statistical test, such as rejection of the null that the indirect effect is zero, or a confidence interval that does not overlap zero. Inference about a single indirect effect is not the purpose of this paper. Suffice it to say that the approach most widely recommended and used of the many that have been proposed is the bootstrap confidence interval in one of several forms (see, e.g., Hayes & Scharkow, 2013; MacKinnon et al., 2004; Shrout & Bolger, 2002), a method that will be emphasized in this paper as a means of comparing indirect effects in models with more than one mediator. The mechanics of bootstrapping and confidence interval construction are described later.

Although the simple mediation model (i.e., a model with a single X, M, and Y) is the most popularly estimated mediation model in the research literature, it is unlikely that something as complex as our thoughts, feelings, and actions can be adequately described by such a simple three-variable causal system. A multiple mediation model is likely a more realistic (albeit still probably oversimplified) model that allows X's effect on Y to be transmitted through more than one mediator simultaneously. An example of a two-mediator parallel multiple mediator model is depicted in Fig. 1B, called a "parallel" model because neither mediator is allowed to affect the other. More complex multiple mediator models are possible, including parallel models with more than two mediators, serial multiple mediators that allow for causal influence between mediators, and blended models with properties of parallel and serial models. For now, we restrict our discussion to the simplest two-mediator parallel case. We address more complex models later.

In the model depicted in Fig. 1B, the direct and two indirect effects of X on Y can be estimated with three equations, one for each of the mediators and one for Y:

$$M_1 = d_{M_1} + a_1 X + \varepsilon_{M_1} \tag{3}$$

$$M_2 = d_{M_2} + a_2 X + \varepsilon_{M_2} \tag{4}$$

$$Y = d_Y + c'X + b_1 M_1 + b_2 M_2 + \varepsilon_Y \tag{5}$$

The indirect effect of X on Y through mediator j is constructed by multiplying a_j and b_j. In this model, there are two indirect effects of X, $a_1 b_1$ and $a_2 b_2$, each called a *specific*

indirect effect. A specific indirect effect is the indirect effect of X on Y through M_j controlling for the other mediator(s). As such, it quantifies something different than if only mediator j is used in a simple mediation model estimated with Eqs. (1) and (2). That is, the two specific indirect effects in a parallel multiple mediator model such as this are likely to be different than the indirect effects estimated in two simple mediation models, each using only one mediator. For simplicity, throughout the rest of this paper, in the context of multiple mediator models, specific indirect effects will simply be referred to as "indirect effects."

Parallel multiple mediation models are commonly estimated by substantive researchers. For example, Wiedow et al. (2013) found that learning as a team significantly improved team outcomes. This relationship was mediated by both increased task knowledge and trust in other teammates. Another study by Copple et al. (2020) explored the relationship between science communication training and scientists' public engagements. They found that a scientist's communication efficacy and attitudes toward the audience mediated this relationship. Rudy et al. (2012) explored whether general self-efficacy and social self-efficacy mediated the relationship between negative self-talk and social anxiety in children. Their analyses only found support for general self-efficacy as a mediator of this relationship. Other examples are abundant in the literature (e.g., Brosowski et al., 2021; Gurmen & Rohner, 2014; Kurti & Dallery, 2014; Singla et al., 2021).

Comparing indirect effects: Value vs. magnitude

Hayes (2022) discusses some of the reasons to estimate a parallel multiple mediator model rather than several simple mediation models each with a single mediator. Most relevant to our discussion going forward is that estimating a multiple mediator model allows for a statistical comparison of indirect effects operating through different mediators. This is possible because any two indirect effects that share X and Y are scaled equivalently (in terms of the metrics of X and Y) even if the mediators are on different scales, allowing them to be formally compared with a statistical test (MacKinnon, 2000; Preacher & Hayes, 2008).

Such statistical comparisons between indirect effects can be useful. For example, an investigator might propose two (or more) processes by which X may affect Y, with the goal of the study to examine not only which mediators are at work but also whether one mechanism is more dominant, stronger, or otherwise different in size than others. Answering this question can be both theoretically and practically important. Consider the case of substance

abuse and intervention research. Mediation analysis is common in these areas of study because it helps identify the mechanisms that lead to behavior change after an intervention (MacKinnon, 2000). If certain mechanisms lead to a greater change in a target behavior than others, then these intervention programs can focus on the more effective pathways and eliminate the ineffective (or less effective) ones, thus improving the quality of treatment and likelihood of behavior change.

Or one theory might predict that *X*'s effect on *Y* is transmitted through mediator A whereas another theory predicts that *X*'s effect on *Y* is carried through mediator B. The goal of the analysis may be to determine which theory has greater support. Perhaps one theory posits that optimism leads to resilience, and this increased resilience is what causes a person to be happy. A different theory might argue that optimism leads to increased self-trust, and this greater self-trust is what causes happiness to increase. If a comparison of the indirect effects suggested that the effect through resilience was larger than the effect through self-trust, the resilience theory about happiness now has more support than the self-trust theory.

Unfortunately, investigators estimating models with more than one mediator often conduct an inferential test for each indirect effect and then fall victim to the fallacy that a difference in significance between effects means they are significantly different (Gelman & Stern, 2006) or that equivalence of significance equals absence of difference. In the former case, if indirect effect A is deemed different from zero but indirect effect B is not, the researcher then concludes that indirect effect A is larger or more important in explaining the relationship than indirect effect B. But knowing that one indirect effect is statistically different from zero while another is not does not mean they are different from each other. In the latter case, knowing that two indirect effects are both different from zero doesn't mean they are equal in value. A variant of this scenario is to conclude when neither indirect effect is different from zero that both mediators are equally uninformative about mechanisms at work transmitting *X*'s effect on *Y*. But it is possible that two effects that are *both* deemed not different from zero by a formal inferential test could still be different from each other when compared with an inferential test. In short, conducting an inference about the size of two indirect effects doesn't provide an inference about the size of the difference between them. A formal statistical comparison of two indirect effects is required to claim they are the same or different.

Fortunately, some investigators do take the additional and, in our opinion, required step of comparing indirect effects to each other before making claims about their relative sizes. This is typically done by calculating what we will henceforth call the *raw difference* between the two indirect effects,

$$a_1b_1 - a_2b_2 \qquad (6)$$

and then conducting an inference about this difference. For instance, in the example of optimism and happiness above, one could compare the indirect effects through resilience (a_1b_1) and self-trust (a_2b_2) by calculating their raw difference and using one of the inferential tests discussed in the next section to determine whether this difference is significantly different from zero. In a real-world example, Scogin et al. (2015) reported the results of a multiple mediator model showing that even though the indirect effect of emotional distress on quality of life through hopelessness was over three times larger than the indirect effect through engagement in pleasant events ($a_1b_1 = -0.13$, $a_2b_2 = -0.04$, leading to a difference of -0.09 using the expression in (6)), a comparison of the two indirect effects using the raw difference revealed they were not significantly different from each other and, consequently, neither could be deemed a stronger mediator than the other. Other examples of the application of the raw difference to compare indirect effects include: Grund and Fries (2014), Romero-Moreno et al. (2016), Schotanus-Dijkstra et al. (2019), and Yıldız (2016).

Conducting inference for the difference

MacKinnon (2000) discussed inference for the raw difference between two indirect effects using the multivariate delta method. This method assumes the sampling distribution of the difference is normal in form. With this assumption, the estimate of the difference can be divided by an estimate of its standard error, producing a test statistic whose *p*-value for testing the null hypothesis of no difference can be derived using the standard normal distribution. Alternatively, a normal-theory confidence interval can be constructed as the point estimate ± 1.96 standard errors. The formula for the standard error of a difference between indirect effects is quite complex and its accuracy as an estimator requires meeting many assumptions.

However, assumptions don't need to be made about the sampling distribution of the difference to compare two indirect effects. A bootstrap confidence interval can be used instead (Preacher & Hayes, 2008). To bootstrap the distribution of the raw difference, draw *B* random samples (where *B* is preferably at least 5000) of size *n* from the data with replacement, where *n* is the original sample size. In each of these *B* bootstrap samples, estimate the model and calculate $a_1b_1 - a_2b_2$. To construct a percentile bootstrap confidence interval for inference, the lower and upper bounds of a *c*% confidence interval using the percentile method are the [(100 − *c*)/2]th and [(*c*/2) + 50]th percentiles of the sorted distribution of *B* bootstrap estimates, respectively. Evidence of a difference between the indirect effects is found in a confidence

interval that does not include zero.[1] This method makes no assumptions about the shape of the sampling distribution of the difference, and because it does not require an estimate of the standard error of the difference, it is easy to implement without making the assumptions typically required for unbiased estimation of the standard errors of estimates.

Comparing opposing indirect effects

Using the raw difference as an estimator of the difference between two indirect effects leads to an inference about whether the two indirect effects are equal in *value*—meaning whether the indirect effects are the same number or different numbers. This is not necessarily the same as an inference about whether the indirect effects are equal in *magnitude*. The magnitude of an indirect effect is its distance from zero and is what many people would think of as the size of the indirect effect or its *strength*. Two indirect effects could be different in value but equal in magnitude (or strength). Two things that are different in magnitude are, necessarily, different in value. But two things that are different in value are not necessarily different in magnitude. That means that an inference that two indirect effects are different in value using the raw difference approach described above does not necessarily lead to the inference that the indirect effects are different in magnitude or strength. The raw difference approach quantifies difference in magnitude *only for indirect effects that are the same sign*.

To illustrate this distinction between value and magnitude, consider an intervention to reduce alcohol consumption among alcoholics. Various mechanisms to reduce alcohol use might be targeted by such an intervention, including increasing self-efficacy and reducing depression (O'Rourke & MacKinnon, 2018). However, intervention programs can have unintended effects on behavior that work in opposition to the targeted mechanisms of the invention. For example, a physical activity intervention might have certain beneficial effects on mental health but could also increase body mass because of higher caloric intake (Cerin & MacKinnon, 2009). And drug-prevention programs could (unintentionally) inform participants of more reasons to use a drug (MacKinnon, 2000). Such cases of opposing indirect effects are an example of *inconsistent mediation* in the context of multiple mediator models (Davis, 1985).[2]

Assume self-efficacy and reasons to drink were measured as part of an evaluation of an intervention to reduce alcohol

use. When looking at the effect through self-efficacy, the indirect effect (a_1b_1) could be negative as expected, with the intervention increasing self-efficacy ($a_1 > 0$) which in turn decreases drinking behavior ($b_1 < 0$), and so $a_1b_1 < 0$. But the intervention might also remind participants of additional reasons to drink ($a_2 > 0$) which in turn increases alcohol use ($b_2 > 0$), resulting in a positive indirect effect through reasons to drink ($a_2b_2 > 0$). Both mechanisms could be at work transmitting the effect of the intervention, but they operate in different directions.

Now further suppose that the indirect effect through self-efficacy, a_1b_1, was –0.60 and the effect through reasons to drink, a_2b_2, was 0.60. The raw difference between them is not zero ($a_1b_1 - a_2b_2 = -1.20$), and an inference that relies on the raw difference could very well lead to the correct inference that the indirect effects are not the same *value*. But it is not as if one of these is necessarily more important than the other in explaining drinking behavior following the intervention. They are equal in *magnitude* or *strength* because they are equidistant from zero—they are merely operating in different directions. Put simply, the mechanism through self-efficacy reduces drinking by the same amount that the mechanism through reasons to drink increases drinking.

Perhaps this example is far-fetched even if it does make the point. But findings like these do exist in the literature. For instance, Chen et al. (2017) found that employee mentoring (*X*) may reduce work–family conflict (*Y*) by increasing access to job-related resources (M_1) which in turn can lower work–family conflict ($a_1b_1 < 0$). At the same time, that mentoring can increase feelings of workload (M_2), which results in more work–family conflict ($a_2b_2 > 0$). Some other examples of opposing indirect effects in parallel multiple mediation models include Levant and Wimer (2014); Pitts et al. (2018); Romano and Balliet, 2017; and Ter Hoeven et al. (2016).

So, if one wants to test whether two indirect effects are different in magnitude when they are of different signs, a different method must be used. We offer some approaches below that can be used to test whether two indirect effects differ in magnitude that also, by definition, double as tests of difference in value. All but one of these can be used regardless of the whether the signs of the indirect effects are the same or different.

The difference in the absolute values One possibility is using the difference between the absolute values of the indirect effects,

$$|a_1b_1| - |a_2b_2| \tag{7}$$

This difference would be zero in a scenario like described above (i.e., when the two opposing indirect effects are equal in magnitude but opposite in sign). Bootstrap confidence intervals can be constructed for inference about this difference as

---

[1] Bias correction of the endpoints can be applied if desired, using computations described in Efron and Tibshirani (1993).

[2] In simple mediation models, inconsistent mediation is present when the indirect effect and direct effect have different signs. Inconsistent mediation is present in multiple mediator models when one or more indirect/direct effects differ in sign.

described above. If the confidence interval for the difference does not contain zero, then the two indirect effects are deemed different in magnitude (and value), regardless of whether they are of the same or different sign.

The sum Ter Hoeven et al. (2016) had the insight that two opposing indirect effects can be compared by using their *sum*

$$a_1b_1 + a_2b_2 \tag{8}$$

In a parallel multiple mediator model with two mediators, this sum is equivalent to the total indirect effect (which is not usually interpreted as a quantification of the difference in magnitude of the two indirect effects). However, when indirect effects are of opposing signs and equal magnitude, the expected value of the sum is equal to zero. The standard error of the sum is available and easy to calculate, though derivation of a p value or confidence interval using the standard error requires assuming the sum is normally distributed. Inference through bootstrapping doesn't require this assumption. If a bootstrap confidence interval doesn't contain zero, then the indirect effects are deemed different in magnitude (and value). It is worth emphasizing that if the two indirect effects have the same sign, this approach provides *no* information about difference in magnitude *or* value.

The ratio A comparison of indirect effects does not require the use of a difference or sum. The ratio of the indirect effects can also be used, though specific indirect effects that are zero or near zero can produce problems with ratio-based measures, as we discuss below. To test whether two indirect effects are different using the simplest ratio approach, we calculate

$$\frac{a_1b_1}{a_2b_2} \tag{9}$$

Like the difference between absolute values, a bootstrap confidence interval for this ratio can be used for inference about both magnitude and value regardless of the signs of the two indirect effects. When the two indirect effects are the same in value (both positive or both negative), this ratio is one, whereas when they are equal in magnitude but opposite in sign, the ratio is negative one. A bootstrap confidence interval that doesn't include one or negative one supports the claim that the two indirect effects are different in magnitude (and value).

A concern with this and some related approaches discussed next is the presence of a zero in the denominator, which produces an indeterminate statistic. In (9), if a_2b_2 is zero, then the result is indeterminate and impossible to interpret. And if a_1b_1 is zero and a_2b_2 is nonzero, the ratio is 0. But it is arbitrary which indirect effect goes in the numerator and which in the denominator. Whether the ratio is zero or

indeterminant in these cases is determined by an arbitrary computational decision. Although this is true, it is extremely rare to observe or compute a value that is exactly zero. Even still, if one of the indirect effects (the point estimate or in a bootstrap sample) is *close* to zero, the ratio can be very small or very large. Bootstrap distributions that are very unusual in form and untrustworthy for construction of confidence intervals could result whenever one of the indirect effects is very small.

The ratio of absolute values The point estimate that results from the ratio of the indirect effects can range from $(-\infty, \infty)$. A modification is the ratio of the *absolute values* of the indirect effects:

$$\frac{|\,a_1b_1\,|}{|\,a_2b_2\,|} \tag{10}$$

This ratio is bounded from $[0, \infty)$. A bootstrap confidence interval that does not contain one results in the inference that the indirect effects are different in magnitude (and value) regardless of their signs. But the presence of a zero or near-zero value in the numerator or denominator may pose the same problems as it did with the simple ratio of the indirect effects.

The proportional absolute value It is possible to further bound the ratio of indirect effects. The ratio of the absolute value of one indirect effect to the sum of the absolute values of the two indirect effects, which we will call the *proportional absolute value*,

$$\frac{|\,a_1b_1\,|}{|a_1b_1|+\,|\,a_2b_2\,|} \tag{11}$$

is bound between $[0, 1]$. It would be equal to 0.5 if the two indirect effects are equal in magnitude (and value). Thus, a bootstrap confidence interval that does not contain 0.5 provides evidence that the two indirect effects are different in magnitude (and value) regardless of their signs. But as with the other ratios, the meaningfulness of this statistic is contingent on neither the numerator nor denominator equaling zero.

Note that inference using the ratio of absolute values (10) should generally produce an identical inference as the use of the proportional absolute value (11), as these are just monotonic transformations of each other. To demonstrate why this is, say $|a_1b_1|$ is 8 and $|a_2b_2|$ is 2. Plugging these values into (10) produces a value of 4 whereas (11) produces a value of .8. However, if we take $\frac{.8}{(1-.8)}$ we get back to the value from Eq. (10) (i.e., 4). This is true for any two values you plug into the equations except when a_2b_2 is zero, which is indeterminate for (10) but 1 for (11). We argue the proportional absolute value produces a more elegant estimate than the

ratio of absolute values that is easier to interpret (the values approach zero and one instead of zero and infinity), and since they provide the same information about relative size of the two indirect effects, we recommend Eq. (11) over Eq. (10) if one of these is to be used.

Example application

We next illustrate the use of these approaches in comparing indirect effects. The health of veterans returning from war has been of keen interest to researchers, citizens, and of course, veterans and active military personnel. It is well understood that combat experiences are a cause of posttraumatic stress in veterans. Understanding the process(es) by which combat experiences lead to posttraumatic stress and other negative outcomes like alcohol abuse and suicide is of the utmost importance for those wishing to better the lives of veterans who have given a part of theirs to help protect the country.

In a retrospective study, Pitts et al. (2018) explored a few mechanisms by which veterans' combat experiences (X) could cause posttraumatic stress (PTS) symptoms (Y) once they returned home from war. They hypothesized that army medics who experienced combat during their most recent deployment would perceive more threats to their life during the deployment (M_1), and this greater level of perceived threat would lead to more PTS symptoms upon returning home. They also hypothesized that combat experience would lead to greater perceived benefits of deployment (M_2), which would in turn lead to fewer PTS symptoms. They fit a parallel multiple mediator model as in Fig. 1B, with socially desirable response tendencies used as a covariate by adding it to the equations for both mediators and the outcome, Y.

The data from the original study come from a questionnaire administered to 324 Army combat medics. As the data are only correlational in nature, causality cannot be unequivocally established, but it is still possible to estimate the indirect effects of interest assuming causality. Without access to their data—and only for the purpose of illustration—we used their published matrix of correlations between variables as well as their means and standard deviations to produce a simulated data set that exactly reproduced these statistics (and it is this simulated dataset that we analyzed). The point estimates of effects we report in our example as well as the substantive interpretation of our analysis reproduce theirs—except they did not conduct the comparisons between indirect effects using the methods we employ in this example. Had they done so, their results using their data might be different even if the substantive conclusions are the same. (Due to the random nature of bootstrapping, the distribution of effects using different data sets would likely produce somewhat different results even if the models, means,

standard deviations, and correlations between variables in the model are the same.) The analysis was conducted using the PROCESS macro for SPSS, SAS, and R (Hayes, 2022).[3] In the Appendix, we provide PROCESS code and output and follow-up SPSS, SAS, and R code for processing the bootstrap estimates from PROCESS, as well as code for conducting the analysis using the *lavaan* package in R.

Estimating Eqs. (3–5) (adding socially desirable responding to each equation as Pitts et al., 2018, did) yields the following results (also summarized in Table 1). More combat experiences were associated with greater perceived threat to life ($a_1 = 0.691$) and greater perceived benefits of deployment ($a_2 = 0.190$). As expected, controlling for other variables in the model of PTS symptoms, perceiving more threats to life was associated with more PTS symptoms ($b_1 = 0.235$), and the more benefits one perceived from deployment, the fewer PTS symptoms they reported ($b_2 = -0.378$). The specific indirect effects of combat experience on PTS symptoms are constructed by multiplying the a and b paths, and inference for each indirect effect was conducted using a percentile bootstrap confidence interval based on 10,000 bootstrap samples. The (specific) indirect effect of combat experiences through perceived threat is $a_1b_1 = 0.162$, 95% CI = [0.057, 0.277]. Assuming causality, a one-unit increase in combat experience seems to *increase* PTS symptoms indirectly (because a_1b_1 is positive) by 0.162 units by increasing feelings of threat to life (because a_1 is positive) which in turn increases PTS symptoms (because b_1 is positive). The specific indirect effect through perceived benefits of deployment is $a_2b_2 = -0.072$, 95% CI = [−0.147, −0.018]. Again, assuming causality, a one-unit increase in combat experience seems to decrease PTS symptoms indirectly (because a_2b_2 is negative) by 0.072 units by first increasing perceived benefits of deployment (because a_2 is positive) which in turn decreases PTS symptoms (because b_2 is negative). As both confidence intervals for the indirect effects exclude zero, we can say that the effect of combat experiences on PTS symptoms operates through both mechanisms *simultaneously* to cause a decrease (and increase) in PTS symptoms. These mechanisms are working in opposition to each other.

But are these indirect effects the same or different in value? In magnitude? We used all methods discussed above to compare a_1b_1 and a_2b_2. The confidence intervals and standard deviations of the bootstrap distributions can be found in Table 1, and Fig. 2 visualizes the bootstrap distributions for these metrices of difference. The confidence interval for the raw difference leads to the conclusion that the indirect effects are different in value ($a_1b_1 - a_2b_2 = 0.234$; 95% CI = [0.118, 0.359]). But because the indirect effects

---

[3] Using an SEM program capable of the computations we have discussed will produce identical results.

Table 1 Unstandardized model coefficients, indirect effects, and comparisons between indirect effects from the analysis of the simulated Pitts et al. (2018) data

	Threat to life (M_1)		View of deployment (M_2)		Post-traumatic stress symptoms (Y)	
Combat (X)	$a_1 \rightarrow$	0.691	$a_2 \rightarrow$	0.190	$c' \rightarrow$	0.335
Threat to life (M_1)					$b_1 \rightarrow$	0.235
View of deployment (M_2)					$b_2 \rightarrow$	–0.378
Social desirability		– 0.519		0.724		–1.523
	Value		Boot SE		95% Boot CI	
$a_1 b_1$	0.162		0.056		(0.057, 0.277)	
$a_2 b_2$	–0.072		0.033		(–0.147, –0.018)	
$a_1 b_1 - a_2 b_2$	0.234		0.061		(0.118, 0.359)	
$|a_1 b_1| - |a_2 b_2|$	0.090		0.068		(–0.046, 0.226)	
$a_1 b_1 + a_2 b_2$	0.090		0.068		(–0.046, 0.226)	
$a_1 b_1 / a_2 b_2$	–2.260		7.992		(–9.969, –0.590)	
$|a_1 b_1| / |a_2 b_2|$	2.260		7.926		(0.602, 10.092)	
$|a_1 b_1| / (|a_1 b_1| + |a_2 b_2|)$	0.693		0.137		(0.376, 0.910)	

Boot CI = percentile bootstrap confidence interval; Boot SE = bootstrap estimate of standard error (the standard deviation of the 10,000 bootstrap estimates)

are opposite in sign, we cannot conclude with this result that they are different in magnitude or strength. Using the difference in absolute values—which requires the confidence interval not include zero to claim differences in magnitude—does not support a claim of difference in magnitude, $|a_1 b_1| - |a_2 b_2| = 0.090$, 95% CI = [−0.046 to 0.226]. Nor does a confidence interval based on the sum of the indirect effects support such a claim, $a_1 b_1 + a_2 b_2 = 0.090$, 95% CI = [−0.046, 0.226]. The other three methods for comparing magnitude that involve division produce the same conclusion: No difference in the magnitude of the two indirect effects. The bootstrap confidence interval for the ratio of indirect effects included −1, $a_1 b_1 / a_2 b_2 = -2.250$, 95% CI = [−9.969, −0.590], and the interval for the ratio of absolute values included 1, $|a_1 b_1| / |a_2 b_2| = 2.250$, 95% CI = [0.602, 10.092]. A claim of difference in magnitude using the proportional absolute value requires the confidence interval exclude 0.50. That criterion was not met: $|a_1 b_1| / (|a_1 b_1| + |a_2 b_2|) = 0.692$, 95% CI = [0.376, 0.910].

Notice that, in this example, the point estimates and confidence intervals for the difference in absolute values and sum approaches are identical. This would tend to occur in larger sample sizes when the point estimates of the indirect effects are sufficiently distant from zero such that rarely (if ever) are any of the bootstrap estimates of the indirect effects different in sign than the point estimates. In such a case, the sum and the difference in absolute values will have essentially the same bootstrap distribution. This is not a general phenomenon but, rather, specific to this example.

It is also important to point out the existence of many extreme bootstrap estimates of the ratio and ratio of absolute value bootstrap distributions noted in two of the panels in Fig. 2. As mentioned earlier, these two statistics can explode in value toward positive or negative infinity as the denominator approaches zero. A few dozen estimates of these two quantities in our analysis do not show up in the histograms because scaling the horizontal axis to include them obscures the appearance of the distribution where most of the bootstrap estimates reside. The consequence of these extremes is a standard error for the statistic that is huge relative to the point estimates, as can be seen in Table 1. This would tend to occur frequently when one or more of the indirect effects is close to zero. Although standard errors are not used in the derivation of the percentile bootstrap confidence interval, many extremes in the bootstrap distribution (a few bootstrap estimates were greater than 400 or less than −400) can influence the width of confidence intervals and this no doubt can affect performance of these approaches in some circumstances. Indeed, notice the very wide width of the confidence intervals (from Table 1) for these two metrics of difference.

This analysis is consistent with the claim that both mechanisms by which combat experience can influence posttraumatic stress appear to be operating, but they do so in different directions. Furthermore, the indirect effect of combat experience on posttraumatic stress through perceived benefits is indeed different in *value* from the indirect effect through perceived threats to life. However, it would be inappropriate and incorrect to say that one of the indirect effects is stronger or weaker than the other. Using the metrics of difference in magnitude we discuss here, the proper conclusion is that neither mechanism is stronger or larger in magnitude than the other.
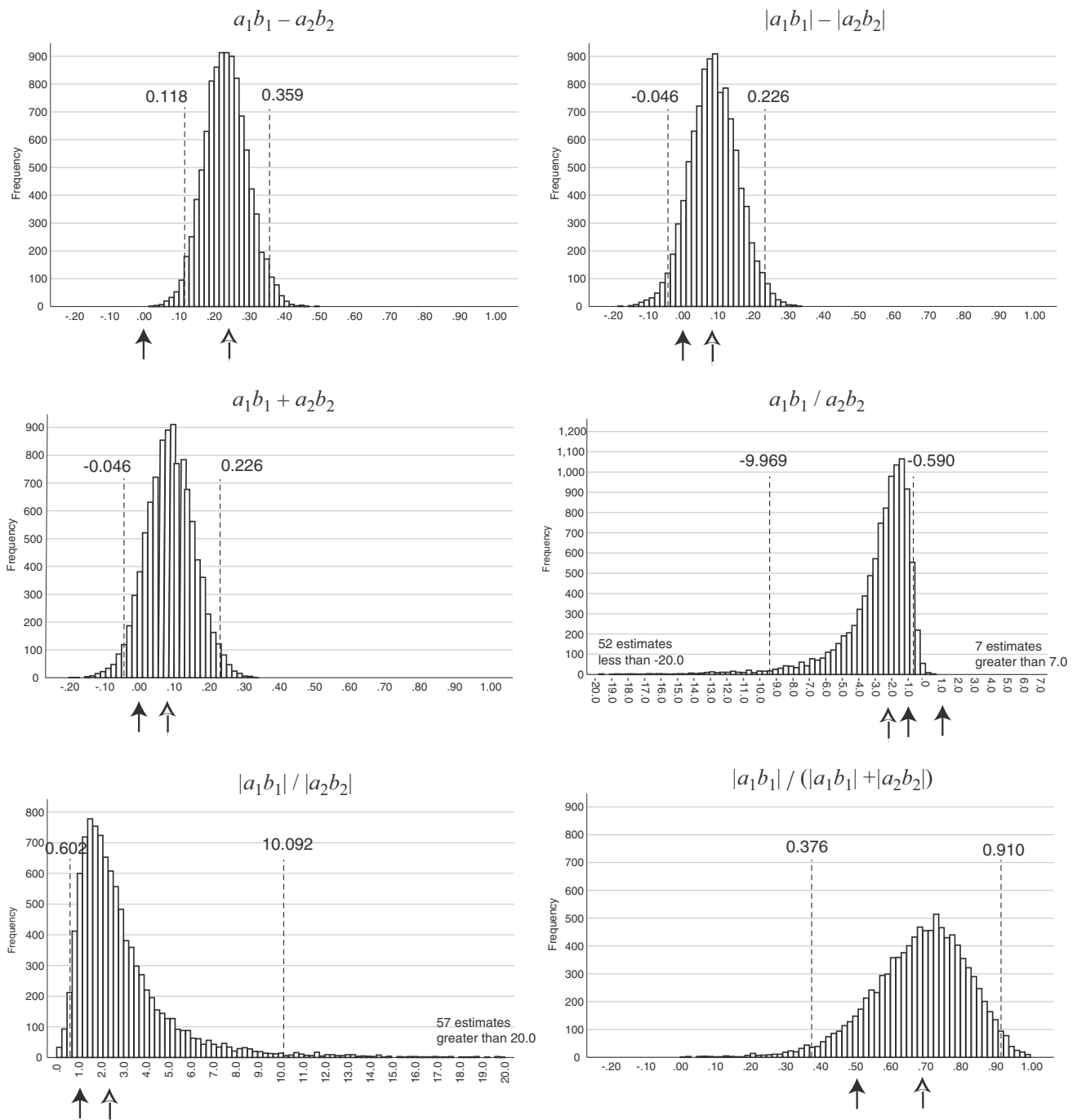
Fig. 2 Bootstrap distributions of the difference between indirect effects from the analysis of the Pitts et al. (2018) simulated data, with 95% bootstrap confidence interval endpoints next to *dashed lines*. The *solid arrow* points at the test value that confidence interval should exclude to conclude two indirect effects are different. The *hollow arrow* points at the point estimate of the difference

Although the raw difference is the predominant method of comparing indirect effects (especially indirect effects of consistent signs), no formal study has evaluated its efficacy. Additionally, as shown in the previous example, the raw difference does not function as a test of equality of magnitude when the indirect effects have opposing signs. Most of the methods we proposed above serve as a test of equality of magnitude *regardless* of the consistency of the signs of the indirect effects. Thus, we aimed to evaluate the performance of the raw difference (the current standard) relative to the methods we proposed when the indirect effects have the same sign and evaluate the performance of the proposed

methods when the current standard is inappropriate as a test of equality of magnitude—that is, when the indirect effects are of different signs.

Relative performance: Simulation findings and recommendations

With several plausible metrics of difference that can be used to compare two indirect effects, it is important to answer the question: Does one perform better than another or is the choice of which to use largely inconsequential? To help answer this question, we conducted a Monte Carlo simulation, the results of which we only briefly summarize here with an eye toward practical recommendations. The simulation examined the performance of these methods for comparing the magnitude of two indirect effects in a parallel multiple mediator model with two mediators. The simulation used a variety of sample sizes (20, 50, 100, 200, 500) and combinations of the values of the *a* and *b* paths (−0.59, −0.39, −0.14, 0, .14, .39, .59), with normal population distributions of *X* and standard normal errors in estimation of *M* and *Y*, as is typical in simulations examining the performance of tests of mediation (e.g., Fritz & MacKinnon, 2007; Hayes & Scharkow, 2013). We tested the null hypothesis of equality of magnitude using 95% bootstrap confidence intervals, rejecting the null when the confidence interval excluded the pertinent test value (e.g., 0 for the raw difference). We also quantified confidence interval coverage, meaning how frequently the confidence interval included the true difference in magnitude. We summarize the simulation findings in terms of type I error, power, and confidence interval coverage before offering recommendations about which method to prefer depending on whether the indirect effects are the same or different in sign. For all the details of the simulation and a more comprehensive discussion of the results, see Coutts (2020).

Type I error and power

All methods performed reasonably well across most conditions. Although all methods were conservative (excessively low type I error rates) in smaller samples, false positives converged upward to the nominal rate of 0.05 as the sample and effect size increased. The metrics based on absolute values (i.e., difference and ratio of absolute values, proportional absolute value) were usually more conservative than the raw difference or sum approaches. Therefore, (and not surprisingly) these absolute-value based measures were correspondingly lower in power than the raw or sum metrics. However, in some conditions, the absolute value-based metrics had excessively high type I error rates—as high as

10–14%. This generally occurred when both indirect effects were zero, either because all paths were zero or one of the paths was nonzero and large but the other paths were zero. Put simply, the raw difference had slightly closer-to-nominal type I error (and higher power) when the indirect effects were of the same sign, and the sum had similarly better performance when the indirect effects were of different signs.

Coverage of the true difference

Although type I error rate and power (1 – type II error rate) are important to consider when choosing a test, coverage of the true value is also an important indicator of the quality of confidence interval-based methods. Looking at the frequency of coverage of the true value, performance mirrored the type I error and power rate results discussed above, with the raw difference and sum approaches performing better than the other methods (for indirect effects of the same and different signs, respectively).

A seemingly anomalous finding was that coverage of the true contrast value was *zero* for the ratio of absolute values and proportional absolute value methods when one of the indirect effects was zero. To understand why this happens, consider the case where the population values of a_1b_1 and a_2b_2 are different but one is zero (e.g., $a_1b_1 = 0$, $a_2b_2 = -0.3$). In this scenario, the population ratio is zero for these metrics. But the only way a bootstrap estimate of this ratio could be as small as zero (and remember that neither of these metrics can be less than zero) is if a_1b_1 in a bootstrap sample is *exactly* zero, which for all intents and purposes won't ever happen. Consequently, a bootstrap confidence interval would never contain the population value in this scenario.[4] Because the population contrast value is unlikely to be covered by the confidence intervals for the division-based absolute value methods when one of the population indirect effects is zero, we recommend avoiding these methods (since other methods perform as well as or better without this problem).

Summary and recommendations

Due to the slightly superior coverage, these results lead us to recommend the use of the raw difference or sum when the interest is in testing the equality of magnitude of two indirect effects (see Table 2 for a summary of our recommendations). But because the raw difference is sensitive to difference in

---

[4] The placement of the indirect effects in the equation is arbitrary, but if they were flipped, the same problem would arise as the population contrast value would be 1 for the proportional absolute value or indeterminate for the ratio of absolute values—both are which are as equally unlikely to occur in an observed or bootstrap sample as 0.

Table 2 A summary of the recommendations from the simulation study when the question is of equality of magnitude

Method	Test of equality of magnitude?	Limitation(s)	Recommended?
Raw difference	Yes, when indirect effects are of the same sign	Not a universal test of equality of magnitude	With *a priori* hypothesis
Sum	Yes, when indirect effects are of opposing signs	Not a universal test of equality of magnitude	With *a priori* hypothesis
Absolute difference	Yes	Slightly lower Type I error and power than raw difference and sum	For exploratory hypotheses or with limited substantive theory
Ratio	Yes	Problems with zero or near-zero values	No
Ratio of absolute values	Yes	Problems with zero or near-zero values	No
Proportional absolute value	Yes	Problems with zero or near-zero values	No

As effect size and sample size increase, these methods converge in performance

magnitude only when the indirect effects are like-signed, and the sum only when the indirect effects are opposite in sign, we suggest that an a priori hypothesis should guide the choice. When theory or other reasoned arguments predict indirect effects of opposing signs, use the sum of the indirect effects to compare their magnitude, whereas the raw difference should be used when the indirect effects are predicted to be of the same sign.

But there may be circumstances in which there are no a priori expectations for the signs of indirect effects, or an investigator is merely exploring differences between indirect effects without a guiding theoretical framework. In that case, we recommend using the difference in absolute values as the metric, as it is sensitive to difference in magnitude irrespective of the signs of the indirect effects. Although it is slightly conservative in smaller samples and has excessive type I error rates in some situations, many would find this worth the price for a more general test. Although the other absolute-value based measures are mathematically equivalent and will lead to the same decision in many circumstances, we have trouble recommending a confidence interval-based inferential tool when coverage is necessarily zero in certain situations or when the population value of the statistic can be indeterminant (as for the ratio and ratio of absolute value metrics).

Discussion

Indirect effects provide information relevant to understanding the mechanism(s) by which some causal effect of X on Y operates. Given that most effects probably operate through multiple mechanisms simultaneously, it makes empirical sense for a mediation model to capture those mechanisms at once in a multiple mediator model. Doing so affords the ability not only to estimate different mechanisms at work but also to determine and test hypotheses about the relative magnitude or strength of those mechanisms in driving the effect. This is accomplished by comparing, descriptively and inferentially, two indirect effects operating through different mediators.

Though still relatively rare in the literature, researchers do sometimes formally compare two indirect effects, typically using the raw difference. When the indirect effects are of the same sign, this seems to be a reasonable approach. However, when the indirect effects being compared differ in sign, which occurs when two mechanisms compete against each other in the direction they "move" Y as a function of $X,$ the raw difference between indirect effects only tests whether those effects differ in value. Thus, one would need to use a different contrast metric to test equality of strength. From our simulation results, we recommend using either the sum of the indirect effects or the difference in their absolute values as the metric for comparing two opposing indirect effects.

Our focus in this paper has been on comparing two specific indirect effects in a mediation model with multiple mediators. For simplicity, we have couched our discussion in the context of the most basic of such models with two mediators operating in parallel with all variables observed rather than latent. In such a model, only one comparison between indirect effects is possible. But our argument for the need to compare indirect effects and the methods we describe for doing so generalize to models with any number of mediators working in parallel or serial as well as latent variable models. In a parallel multiple mediator model with k mediators, there are $0.5(k^2 - k)$ possible comparisons between specific indirect effects, and in a serial model that includes all possible indirect effects, there are $k^2 - k - 1$ possible comparisons. For some of these comparisons, the raw difference approach might be appropriate, but for others, a method that is sensitive to questions about difference in magnitude for indirect effects that differ in sign may be required to test the hypothesis of interest. A tool like PROCESS could not be used for latent variable models, but SEM programs are built for this and are widely available and relatively easy to use with a

bit of training (for a discussion of latent variable mediation analysis, see Cheung & Lau, 2008; Lau & Cheung, 2012; MacKinnon, 2008).[5] The methods discussed here translate without modification to the comparison of indirect effects defined from the structural part of a latent variable structural equation model.

Our discussion and recommendations from our simulation have been based on the comparison of indirect effects in their unstandardized form. In this form, an indirect effect scales differences in Y due to a one-unit difference in X resulting from X's effect on Y which in turn affects M. The differences in X and Y in the prior sentence are quantified in the original unit of measurement of X and Y. Investigators may prefer to instead scale indirect effects in standard deviations of differences rather than the original metric. Examples of such rescaling, often discussed in the content of measures of "effect size" in mediation analysis, include the partially and completely standardized indirect effects discussed by Preacher and Kelly (2011) and Hayes (2022).

The difference metrics in Eqs. (6–11) can be used for comparing the value or magnitude of partially and completely standardized indirect effects. For the difference between completely standardized indirect effects, multiply each indirect effect by s_X/s_Y prior to the computation of the difference metric, where s_X and s_Y are the standard deviations of X and Y, respectively. For the difference between partially standardized indirect effects, first multiply each indirect effect by $1/s_Y$. For the raw difference (6), difference in absolute values (7) and sum metrics (8), the result is equivalent to multiplying this metric by this same ratio. But when using the ratio (9), absolute value of ratios (10), and proportional absolute value (11) metrics, these rescaling computations are unnecessary, as this rescaling of the specific indirect effects appears in both the numerator and denominator of the difference metric and cancels out. This means that these three metrics of difference can be used unmodified to quantify the difference between unstandardized, partially standardized, or completely standardized indirect effects. And the performance of these methods documented in Coutts (2020) and summarized in this paper generalize exactly.

We believe our recommendations also generalize to the comparison of partially and completely standardized indirect effects when using the raw difference, difference in absolute values, and sum metrics, but we must acknowledge a caveat. Bootstrap estimates of the difference in standardized indirect effects rely on bootstrap estimates of variation in X and Y in each bootstrap sample rather than variation observed in the original sample (cf., Cheung, 2009). This introduces a new source of random variation in bootstrap estimates not present when comparing unstandardized indirect effects. The consequence of this additional source of variation is that the rank order correlation of the bootstrap estimates or distributions of difference using unstandardized, partially standardized, and completely standardized indirect effects may not be one. So when using any of these three metrics of difference, it would be possible for a bootstrap confidence interval for the difference between unstandardized indirect effects to result in a different conclusion than when using standardized indirect effects. However, this would tend to be fairly rare in practice, especially in larger samples. Although this extra source of variation might produce greater conservativism in comparisons based on standardized indirect effects (i.e., reduced power and type I error), we see no basis for believing that any such differences in conclusion would change the relative performance of these metrics to such a degree that our recommendations would change.

Our treatment of the mathematics of mediation analysis has been based on the traditional regression or SEM approach that has dominated the behavioral sciences. The counterfactual or potential outcomes approach is becoming increasingly popular (see e.g., Imai et al., 2010; Pearl, 2012; Vanderweele, 2015; Vanderweele & Vansteelandt, 2014). One of the major differences between the counterfactual approach and the traditional approach is its flexibility in the definition of the indirect effect for different types of models for outcomes (variables on the left-hand sides of the equations) that are dichotomous, ordinal, count, time to event, and so forth that are not as appropriately analyzed using Eqs. (1–5). This counterfactual approach also need not assume the absence of interaction between X and the mediator(s) in the model of Y, as we have in our discussion and mathematical definition of indirect effects. This paper also did not cover more complicated models such as generalized linear models, survival models, or nonlinear models. However, these differences in mathematics and assumptions do not change the generalizability of the methods we discuss for comparing indirect effects (though the simulation results may not generalize beyond what we examined). They merely change the mathematics of the estimation of those indirect effects. Researchers who prefer the counterfactual approach or a different modeling framework to mediation analysis can still use the methods we describe here when their question focuses on whether two indirect effects in a multiple mediator model differ from each other, whether in value or magnitude. Although we suspect our recommendations summarized in Table 2 would not change, it is unknown the extent to which our simulation results would generalize to different modeling frameworks.

---

[5] For some of the methods we discuss, your chosen SEM program must be capable of forming new parameters that are functions of others that include absolute values of structural paths. Not all SEM programs can do this (e.g., Mplus cannot; the laavan package for R can).

Regardless of the scaling of indirect effects or modeling preference, we encourage researchers to move their thinking away from the descriptive and qualitative mindset to comparing indirect effects (e.g., "this effect is significant, that one is not, so these effects differ") in favor of a more rigorous, quantitative orientation that considers the difference between two indirect effects and a hypothesis about that difference as things that can be quantified, estimated, and tested.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13428-022-01988-0.

References

Brosowski, T., Olason, D. T., Turowski, T., & Hayer, T. (2021). The gambling consumption mediation model (GCMM): A multiple mediation approach to estimate the association of particular game types with problem gambling. *Journal of Gambling Studies, 37*(1), 107–140. https://doi.org/10.1007/s10899-020-09928-3

Cerin, E., & MacKinnon, D. P. (2009). A commentary on current practice in mediating variable analyses in behavioural nutrition and physical activity. *Public Health Nutrition, 12*(8), 1182–1188. https://doi.org/10.1017/S1368980008003649

Chen, C., Wen, P., & Hu, C. (2017). Role of formal mentoring in proteges' work-to-family conflict: A double-edged sword. *Journal of Vocational Behavior, 100*, 101–110. https://doi.org/10.1016/j.jvb.2017.03.004

Cheung, G. W., & Lau, R. S. (2008). Testing mediation and suppression effects of latent variables. *Organizational Research Methods, 11*, 296–325. https://doi.org/10.1177/1094428107300343

Cheung, M. W.-L. (2009). Comparison of methods for constructing confidence intervals of standardized indirect effects. *Behavior Research Methods, 41*, 425–438. https://doi.org/10.3758/BRM.41.2.425

Copple, J., Bennett, N., Dudo, A., Moon, W. K., Newman, T. P., Besley, J., et al. (2020). Contribution of training to scientists' public engagement intentions: A test of indirect relationships using parallel multiple mediation. *Science Communication, 42*(4), 508–537. https://doi.org/10.1177/1075547020943594

Coutts, J. J. (2020). *Contrasting Contrasts: An Exploration of Methods for Comparing Indirect Effects in Mediation Models* [Master's thesis, Ohio State University]. OhioLINK Electronic Theses and Dissertations Center. http://rave.ohiolink.edu/etdc/view?acc_num=osu1605579439442685

Davis, J. A. (1985). *The logic of causal order*. SAGE Publications, Inc. https://doi.org/10.4135/9781412986212

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman and Hall.

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological science, 18*(3), 233–239. https://doi.org/10.1111/j.1467-9280.2007.01882.x

Gelman, A., & Stern, H. (2006). The difference between "significant" and "not significant" is not itself statistically significant. *The American Statistician, 60*(4), 328–331. https://doi.org/10.1198/000313006X152649

Grund, A., & Fries, S. (2014). Study and leisure interference as mediators between students' self-control capacities and their domain-specific functioning and general well-being. *Learning and Instruction, 31*, 23–32. https://doi.org/10.1016/j.learninstruc.2013.12.005

Gurmen, M. S., & Rohner, R. P. (2014). Effects of marital distress on Turkish adolescents' psychological adjustment. *Journal of Child and Family Studies, 23*, 1155–1162. https://doi.org/10.1007/s10826-013-9773-7

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (3rd ed.). The Guilford Press.

Hayes, A. F., Montoya, A. K., & Rockwood, N. J. (2017). The analysis of mechanisms and their contingencies: PROCESS versus structural equation modeling. *Australasian Marketing Journal, 25*, 76–81. https://doi.org/10.1016/j.ausmj.2017.02.001

Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science, 24*(10), 1918–1927. https://doi.org/10.1177/0956797613480187

Ho, S. S., Peh, X., & Soh, V. W. (2013). The cognitive mediation model: Factors influencing public knowledge of the H1N1 pandemic and intention to take precautionary behaviors. *Journal of Health Communication, 18*(7), 773–794. https://doi.org/10.1080/10810730.2012.743624

Hoffman, L. H., & Young, D. G. (2011). Satire, punch lines, and the nightly news: Untangling media effects on political participation. *Communication Research Reports, 28*(2), 159–168. https://doi.org/10.1080/08824096.2011.565278

Iacobucci, D., Saldanha, N., & Deng, X. (2007). A mediation on mediation: Evidence that structural equations models perform better than regressions. *Journal of Consumer Psychology, 17*, 140–154. https://doi.org/10.1016/S1057-7408(07)70020-7

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods, 15*, 309–334. https://doi.org/10.1037/a0020761

Jin, M. H., McDonald, B., & Park, J. (2018). Does public service motivation matter in public higher education? Testing the theories of person–organization fit and organizational commitment through a serial multiple mediation model. *The American Review of Public Administration, 48*(1), 82–97. https://doi.org/10.1177/0275074016652243

Kurti, A. N., & Dallery, J. (2014). Effects of exercise on craving and cigarette smoking in the human laboratory. *Addictive Behaviors, 39*, 1131–1137. https://doi.org/10.1016/j.addbeh.2014.03.004

Levant, R. F., & Wimer, D. J. (2014). The Relationship Between Conformity to Masculine Norms and Men's Health Behaviors: Testing a Multiple Mediator Model. *International. Journal of Men's Health, 13*(1). https://doi.org/10.3149/jmh.1301.22

Lau, R. S., & Cheung, G. W. (2012). Estimating and comparing specific mediation effects in complex latent variables. *Organizational Research Methods, 15*, 3–16. https://doi.org/10.1177/1094428110391673

MacKinnon, D. P. (2000). Contrasts in multiple mediator models. In J. Rose, L. Chassin, C. C. Presson, & S. J. Sherman (Eds.), *Multivariate applications in substantive use research: New methods for new questions* (pp. 141–160). Lawrence Erlbaum Associates.

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*(1), 99–128. https://doi.org/10.1207/s15327906mbr3901_4

MacKinnon, D. P. (2008). *Statistical mediation analysis*. Routledge.

O'Rourke, H. P., & MacKinnon, D. P. (2018). Reasons for testing mediation in the absence of an intervention effect: A research imperative in prevention and intervention research. *Journal of Studies on Alcohol and Drugs, 79*(2), 171–181. https://doi.org/10.15288/jsad.2018.79.171

Pearl, J. (2012). The causal mediation formula—A guide to the assessment of pathway and mechanisms. *Prevention Science, 13*, 426–436. https://doi.org/10.1007/s11121-011-0270-1

Pek, J., & Hoyle, R. H. (2016). On the (in)validity of tests of simple mediation: Threats and solutions. *Social and Personality Psychology Compass, 10*, 150–163. https://doi.org/10.1111/spc3.12237

Pitts, B. L., Safer, M. A., Castro-Chapman, P. L., & Russell, D. W. (2018). Retrospective appraisals of threat and benefit mediate the effects of combat experiences on mental health outcomes in army medics. *Military Behavioral Health, 6*(3), 226–233. https://doi.org/10.1080/21635781.2017.1401967

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*, 879–891. https://doi.org/10.3758/BRM.40.3.879

Preacher, K. J., & Kelly, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods, 16*, 93–115. https://doi.org/10.1037/a0022658

Raykov, T., Brennan, M., Reinhardt, J. P., & Horowitz, A. (2008). Comparison of mediated effects: A correlation structure modeling approach. *Structural Equation Modeling, 15*, 603–626. https://doi.org/10.1080/10705510802339015

Rijnhart, J. J., Twisk, J. W., Chinapaw, M. J., de Boer, M. R., & Heymans, M. W. (2017). Comparison of methods for the analysis of relatively simple mediation models. *Contemporary Clinical Trials Communications, 7*, 130–135. https://doi.org/10.1016/j.conctc.2017.06.005

Romano, A., & Balliet, D. (2017). Reciprocity outperforms conformity to promote cooperation. *Psychological Science, 28*, 1490–1502. https://doi.org/10.1177/0956797617714828

Romero-Moreno, R., Losada, A., Márquez-González, M., & Mausbach, B. T. (2016). Stressors and anxiety in dementia caregiving: multiple mediation analysis of rumination, experiential avoidance, and leisure. *International Psychogeriatrics, 28*(11), 1835–1844. https://doi.org/10.1017/S1041610216001009

Rudy, B. M., Davis III, T. E., & Matthews, R. A. (2012). The relationship among self-efficacy, negative self-referent cognitions, and social anxiety in children: A multiple mediator model. *Behavior Therapy, 43*(3), 619–628. https://doi.org/10.1016/j.beth.2011.11.003

Schönfeld, P., Brailovskaia, J., Bieda, A., Zhang, X. C., & Margraf, J. (2016). The effects of daily stress on positive and negative mental health: Mediation through self-efficacy. *International Journal of Clinical and Health Psychology, 16*(1), 1–10. https://doi.org/10.1016/j.ijchp.2015.08.005

Schotanus-Dijkstra, M., Pieterse, M. E., Drossaert, C. H., Walburg, J. A., & Bohlmeijer, E. T. (2019). Possible mechanisms in a multicomponent email guided positive psychology intervention to improve mental well-being, anxiety and depression: A multiple mediation model. *The Journal of Positive Psychology, 14*(2), 141–155. https://doi.org/10.1080/17439760.2017.1388430

Scogin, F., Morthland, M., DiNapoli, E. A., LaRocca, M., & Chaplin, W. (2015). Pleasant events, hopelessness, and quality of life in rural older adults. *Journal of Rural Health, 32*, 102–109. https://doi.org/10.1111/jrh.12130

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*(4), 422–445. https://doi.org/10.1037/1082-989X.7.4.422

Singla, D. R., MacKinnon, D. P., Fuhr, D. C., Sikander, S., Rahman, A., & Patel, V. (2021). Multiple mediation analysis of the peer-delivered Thinking Healthy Programme for perinatal depression: findings from two parallel, randomised controlled trials. *The British Journal of Psychiatry, 218*(3), 143–150. https://doi.org/10.1192/bjp.2019.184

Ter Hoeven, C. L., van Zoonen, W., & Fonner, K. L. (2016). The practical paradox of technology: The influence of communication technology use on employee burnout and engagement. *Communication Monographs, 83*(2), 239–263. https://doi.org/10.1080/03637751.2015.1133920

Van der Heijden, B., Mahoney, C. B., & Xu, Y. (2019). Impact of job demands and resources on Nurses' burnout and occupational turnover intention towards an age-moderated mediation model for the nursing profession. *International Journal of Environmental Research and Public Health, 16*(11), 2011. https://doi.org/10.3390/ijerph16112011

VanderWeele, T., & Vansteelandt, S. (2014). Mediation analysis with multiple mediators. *Epidemiologic Methods, 2*(1), 95–115. https://doi.org/10.1515/em-2012-0010

Vanderweele, T. J. (2015). *Explanation in causal inference: Methods for mediation and interaction*. Oxford University Press.

Wieder, B., & Ossimitz, M. L. (2015). The impact of Business Intelligence on the quality of decision making–a mediation model. *Procedia Computer Science, 64*, 1163–1171. https://doi.org/10.1016/j.procs.2015.08.599

Wiedow, A., Konradt, U., Ellwart, T., & Steenfatt, C. (2013). Direct and indirect effects of team learning on team outcomes: A multiple mediator analysis. *Group Dynamics: Theory, Research, and Practice, 17*(4), 232. https://doi.org/10.1037/a0034149

Yıldız, M. A. (2016). Multiple mediation of emotion regulation strategies in the relationship between loneliness and positivity in adolescents. *Education & Science, 41*(186), 10.15390/EB.2016.6193.

Open Practices Statement

The data for the demonstration are available as a supplementary material and the study was not preregistered.