



Universiteit  
Leiden  
The Netherlands

## How to synthesize randomized controlled trial data with meta-analytic structural equation modeling: a comparison of various d-to-rpb conversions

Jonge, H. de; Kan, K.J.; Oort., F.J.; Jak, S.

### Citation

Jonge, H. de, Kan, K. J., Oort., F. J., & Jak, S. (2025). How to synthesize randomized controlled trial data with meta-analytic structural equation modeling: a comparison of various d-to-rpb conversions. *Psychological Methods*. doi:10.1037/met0000790

Version: Publisher's Version

License: [Licensed under Article 25fa Copyright Act/Law \(Amendment Taverne\)](#)

Downloaded from: <https://hdl.handle.net/1887/4291024>

**Note:** To cite this publication please use the final published version (if applicable).

# Psychological Methods

## **How to Synthesize Randomized Controlled Trial Data With Meta-Analytic Structural Equation Modeling: A Comparison of Various d-to-rpb Conversions**

Hannelies de Jonge, Kees-Jan Kan, Frans J. Oort, and Suzanne Jak

Online First Publication, September 29, 2025. <https://dx.doi.org/10.1037/met0000790>

### CITATION

de Jonge, H., Kan, K.-J., Oort, F. J., & Jak, S. (2025). How to synthesize randomized controlled trial data with meta-analytic structural equation modeling: A comparison of various d-to-rpb conversions. *Psychological Methods*. Advance online publication. <https://dx.doi.org/10.1037/met0000790>

# How to Synthesize Randomized Controlled Trial Data With Meta-Analytic Structural Equation Modeling: A Comparison of Various $d$ -to- $r_{pb}$ Conversions

Hannelies de Jonge<sup>1, 2</sup>, Kees-Jan Kan<sup>1</sup>, Frans J. Oort<sup>1</sup>, and Suzanne Jak<sup>1</sup>

<sup>1</sup> Methods and Statistics, Research Institute of Child Development and Education, University of Amsterdam

<sup>2</sup> Methodology and Statistics, Psychology, Leiden University

## Abstract

Meta-analytic structural equation modeling (MASEM) allows a researcher to simultaneously examine multiple relations among variables by fitting a structural equation model to summary statistics from multiple studies. Consider, for example, a mediation model with a predictor ( $X$ ), mediator ( $M$ ), and outcome variable ( $Y$ ). In such a model,  $X$  can be a dichotomous variable, allowing researchers to examine the direct and indirect effects of an intervention as in randomized controlled trials (RCTs). However, the natural choice of a meta-analysis of RCTs would involve standardized mean differences as effect sizes, whereas MASEM requires correlation matrices as input. This can be solved by converting standardized mean differences (Cohen's  $d$  or Hedges'  $g$ ) to point-biserial correlations ( $r_{pb}$ ). Possible conversion formulas vary across publications and conversion tools, and it is unclear which one is most appropriate for use in MASEM. The aim of this article is to describe and evaluate several conversions of standardized mean differences to point-biserial correlations in the context of RCTs. We investigate the impact of the usage of various conversions on MASEM parameter estimation using the R package metaSEM in a simulation study, varying the ratio of group sample sizes, number of primary studies, sample sizes, and missingness. The results show that a relatively unknown  $d$ -to- $r_{pb}$  conversion generally performs best. However, this conversion formula is not implemented in the mainstream conversion tools. We developed a user-friendly web application entitled Effect Size Calculator and Converter (<https://hdejonge.shinyapps.io/ESCACO>) that converts the user's primary study statistics into an effect size suitable for use in MASEM.

## Translational Abstract

Meta-analytic structural equation modeling (MASEM) allows a researcher to simultaneously examine multiple relations among variables by fitting a structural equation model to summary statistics from multiple studies. Consider, for example, a mediation model with a predictor ( $X$ ), mediator ( $M$ ), and outcome variable ( $Y$ ). In such a model,  $X$  can be a dichotomous variable, allowing researchers to examine the direct and indirect effects of an intervention as in randomized controlled trials (RCTs). However, the natural choice of a meta-analysis of RCTs would involve standardized mean differences as effect sizes, whereas MASEM requires correlation matrices as input. This can be solved by converting standardized mean differences (Cohen's  $d$  or Hedges'  $g$ ) to point-biserial correlations ( $r_{pb}$ ). Possible conversion formulas vary across publications and conversion tools, and it is unclear which one is most appropriate for use in MASEM. The aim of this article is to describe and evaluate several conversions of standardized mean differences to point-biserial correlations in the context of RCTs. We investigate the impact of the usage of various conversions on MASEM parameter estimation using the R package metaSEM in a simulation study, varying the ratio of group sample sizes, number of primary studies, sample sizes, and missingness. The results show that a relatively unknown  $d$ -to- $r_{pb}$  conversion generally performs best. However, this conversion formula is not implemented in the mainstream conversion tools. We developed a user-friendly web application entitled Effect Size Calculator and Converter (<https://hdejonge.shinyapps.io/ESCACO>) that converts the user's primary study statistics into an effect size suitable for use in MASEM.

Sarah Depaoli served as action editor.

Hannelies de Jonge  <https://orcid.org/0000-0001-5766-8472>

Earlier versions of this article were presented in oral presentations at the International Meeting of the Psychometric Society, European Congress of Methodology, Working Group Structural Equation Modeling, and European Conference on Developmental Psychology. This research was funded by the Dutch Research Council (NWO) for scientific research under the grant VI.Vidi.201.009 awarded to Suzanne Jak. The authors declare that they have no conflicts of interest. The additional online materials are available on the Open Science Framework at <https://osf.io/ynkj3/>.

Hannelies de Jonge served as lead for formal analysis, visualization, data curation, and writing—original draft. Kees-Jan Kan served as lead

for supervision. Suzanne Jak served as lead for funding acquisition. Kees-Jan Kan served in a supporting role for data curation and formal analysis. Kees-Jan Kan and Suzanne Jak served in a supporting role for visualization. Hannelies de Jonge, Kees-Jan Kan, Frans J. Oort, and Suzanne Jak contributed equally to conceptualization, methodology, and validation. Hannelies de Jonge and Kees-Jan Kan contributed equally to software. Frans J. Oort and Suzanne Jak contributed equally to supervision. Kees-Jan Kan, Frans J. Oort, and Suzanne Jak contributed equally to writing—review and editing.

Correspondence concerning this article should be addressed to Hannelies de Jonge, Methodology and Statistics, Psychology, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands. Email: [h.de.jonge@fsw.leidenuniv.nl](mailto:h.de.jonge@fsw.leidenuniv.nl)

**Keywords:** meta-analytic structural equation modeling, randomized controlled trials, standardized mean difference, point-biserial correlation, *d*-to-*r*<sub>pb</sub> conversions

**Supplemental materials:** <https://doi.org/10.1037/met0000790.supp>

Meta-analytic structural equation modeling (MASEM; Becker, 1992, 1995; Cheung, 2015a, 2019; Jak, 2015; Viswesvaran & Ones, 1995) is a statistical technique to synthesize results from multiple independent primary studies. MASEM allows for a systematic investigation of multiple relations among variables simultaneously by fitting a structural equation model (SEM) on a meta-analytic data set. For correlation-based MASEM analysis, such as two-stage structural equation modeling (TSSEM; Cheung, 2014; Cheung & Chan, 2005) or one-stage MASEM (OSMASEM, Jak & Cheung, 2020), a meta-analyst needs to obtain the bivariate correlations between the variables in the hypothesized model from different independent primary studies.

In TSSEM, the correlations from primary studies are first pooled, using full information maximum likelihood as the method of estimation. In the second stage of TSSEM, the SEM is fit to the pooled correlations using weighted least squares estimation. The weighted least squares procedure ensures that larger studies (i.e., studies with larger total sample sizes) weigh more when estimating the parameters in Stage 2. In OSMASEM, the SEM is directly fit to the correlation coefficients using full information maximum likelihood estimation. This estimation method also ensures that larger studies weigh more. An advantage of OSMASEM is that one can include both continuous and categorical moderators, whereas TSSEM only allows for categorical moderators. Both OSMASEM and TSSEM make fixed and random effects modeling possible. In fixed-effects models, observed differences across the primary studies' correlations are solely due to sampling error (Hedges & Vevea, 1998). In random-effects models, one allows for additional heterogeneity in the form of between-studies variance. TSSEM and OSMASEM (random and fixed models) can both be conducted using the R package metaSEM (Cheung, 2015b). One can also apply OSMASEM using a user-friendly R Shiny web application (see Jak et al., 2021). In this article, we focus on random-effects OSMASEM applied to data obtained from (simulated) randomized controlled trials (RCTs).

With RCTs, an advantage of MASEM is that it enables the evaluation of both the direct and indirect effects of the intervention (*X*, a dichotomous variable) on some continuous outcome variable (*Y*) in the presence of a continuous mediating variable (*M*). Even though primary studies may not specifically have investigated mediation, it is possible to investigate this at a meta-analytic level. The focus in the included primary studies may have been only on the bivariate relation between *X* and *M*, between *M* and *Y*, and/or between *X* and *Y*. Meta-analysts may be interested in mediation because this may help explain the underlying mechanisms of the intervention. For example, Jansen et al. (2019) investigated whether students' engagement in self-regulated learning activities mediated the effect of self-regulated learning interventions on achievement in higher education. Ng et al. (2023) applied MASEM to evaluate mediators of the effect of cognitive behavioral therapy on youth depression.

Correlation-based MASEM has been found to be the most suited for the evaluation of meta-analytic indirect effects (Van Zundert & Miočević, 2020). A complicating factor in such analysis, however, is that articles on RCT typically report standardized mean differences (Cohen's *d* or Hedges' *g*), whereas MASEM requires correlation matrices as input. This can be solved by converting standardized mean differences to point-biserial correlations. Here lies a complication, because in contrast to a standardized mean difference, the point-biserial correlation depends on the distribution of respondents over the control and experimental group. That is, the relative sizes of the groups affect the point-biserial correlation (McGrath & Meyer, 2006). As a result, the same standardized mean difference can lead to different point-biserial correlations. Moreover, conversion formulas differ across publications (see Pustejovsky, 2014), statistical software, and online conversion tools (see Table 1), while it is unclear which conversion is most appropriate to use in MASEM with RCTs. The aim of this article is to describe and evaluate several conversions of standardized mean differences to point-biserial correlations in the context of RCTs. By means of Monte Carlo data simulation, we will investigate the influence of various conversions on the estimation of MASEM parameters, including (pooled) correlations, and the resulting path coefficients and standard errors. To this end, we first review the various options to convert standardized mean differences to point-biserial correlations.

## Standardized Mean Differences

Our starting point for the conversion is the logic behind the RCT, which is as follows (Holland, 1986). Hypothetically, the individuals that comprise the population are simultaneously placed in two situations—intervention and no intervention—so that the difference on the outcome variable is the causal effect of the intervention. The average over individuals would equal the average effect the researcher would be interested in. Since in reality individuals cannot be subjected to two conditions simultaneously—"the fundamental problem of causal inference"—the individuals are randomly divided into two groups. Because of the randomization, there are no systematic differences between the two groups prior to the intervention:  $\mu_1 = \mu_0$  and  $\sigma_1 = \sigma_0 = \sigma$ , where  $\mu$  and  $\sigma$  represent the mean and standard deviation, respectively, with subscript 1 referring to the experimental group receiving intervention, and 0 to the control group receiving no intervention. Except for the intervention itself, the two groups are treated exactly the same. If the intervention is effective, this creates a mean difference  $\delta_\mu = \mu_1 - \mu_0$ . Assuming individual differences within groups are merely noise and do not relate to the intervention,  $\sigma$  remains equal over groups, so that the standardized mean difference on the level of the population can be defined as:

$$\delta = \frac{\mu_1 - \mu_0}{\sigma}. \quad (1)$$

**Table 1***Overview of Statistical Software and Online Tools and Implemented Conversions*

| Statistical software or online tool                                                                                                                                                                                                                                                                                                                                     | Conversion formula (see also Table 2)              |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------|
| R function <code>escalc()</code> within the R package <code>metafor</code> (Viechtbauer, 2010)                                                                                                                                                                                                                                                                          | General conversion (Equation 10) <sup>a</sup>      |
| R function <code>convert_d2r()</code> within the R package <code>esc</code> (Lüdtke, 2019)                                                                                                                                                                                                                                                                              | Large-sample conversion (Equation 11) <sup>b</sup> |
| Practical Meta Analysis Effect Size Calculator – Campbell Collaboration ( <a href="https://www.campbellcollaboration.org/calculator">https://www.campbellcollaboration.org/calculator</a> ; The subtab “correlation” under the tab “...Means and standard deviations”)                                                                                                  | Large-sample conversion (Equation 11) <sup>c</sup> |
| Effect Size Determination Program (Lipsey & Wilson, 2001) in Microsoft Excel (tab: Means and SDs) <sup>d</sup>                                                                                                                                                                                                                                                          | Large-sample-equal-groups conversion (Equation 13) |
| R function <code>d_to_r()</code> within the R package <code>effectsize</code> (Ben-Shachar et al., 2020)                                                                                                                                                                                                                                                                | General conversion (Equation 10) <sup>e</sup>      |
| <a href="https://www.escal.site/">https://www.escal.site/</a>                                                                                                                                                                                                                                                                                                           | Large-sample-equal-groups conversion (Equation 13) |
| <a href="https://www.psychometrica.de/effect_size.html#cohen">https://www.psychometrica.de/effect_size.html#cohen</a> (tab 14)                                                                                                                                                                                                                                          | Large-sample-equal-groups conversion (Equation 13) |
| <a href="https://view.officeapps.live.com/op/view.aspx?src=http%3A%2F%2Fwww.stat-help.com%2Fspreadsheets%2FConverting%2520effect%2520sizes%25202012-06-19.xls&amp;wdOrigin=BROWSELINK">https://view.officeapps.live.com/op/view.aspx?src=http%3A%2F%2Fwww.stat-help.com%2Fspreadsheets%2FConverting%2520effect%2520sizes%25202012-06-19.xls&amp;wdOrigin=BROWSELINK</a> | Large-sample-equal-groups conversion (Equation 13) |

*Note.* It is not always directly stated which conversion formula is used. Therefore, we also tested all tools with example data and compared the output with the output of the investigated four conversions.

<sup>a</sup>Based on the information found here: [metafor](https://cran.r-project.org/web/packages/metafor/metafor.pdf): <https://cran.r-project.org/web/packages/metafor/metafor.pdf>. <sup>b</sup>Using Equation 11 but using the notation of Lipsey and Wilson (2001). Based on information found here: <https://cran.r-project.org/web/packages/esc/esc.pdf> and <https://mason.gmu.edu/~dwilsonb/downloads/esformulas.pdf>. <sup>c</sup>Using Equation 11 but using the notation of Lipsey and Wilson (2001). Based on information found here: Practical Meta Analysis Effect Size Calculator – Campbell Collaboration (<https://www.campbellcollaboration.org/calculator>; Subtab “Formulas” under the tab “3.3 Correlation from means and standard deviations”). <sup>d</sup>This is an older Excel-based version of the Practical Meta Analysis Effect Size Calculator – Campbell Collaboration (<https://www.campbellcollaboration.org/calculator>) and can be found at <https://mason.gmu.edu/~dwilsonb/ma.html>. This website is not available anymore. However, in practice, researchers may still have this Excel version on their computer and use it. <sup>e</sup>Based on information from: [https://easystats.github.io/effectsize/reference/d\\_to\\_r.html](https://easystats.github.io/effectsize/reference/d_to_r.html).

Cohen’s  $d$  is a sample estimate of this standardized mean difference in the population. On this (sample) level, the expression of a standardized mean difference is as follows (Borenstein & Hedges, 2019; Borenstein et al., 2009; Lakens, 2013; Lipsey & Wilson, 2001):

$$d = \frac{\bar{y}_1 - \bar{y}_0}{s_p}, \quad (2)$$

with

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_0 - 1)s_0^2}{n_1 + n_0 - 2}}. \quad (3)$$

In these equations, subscripts 1 and 0 again refer to the experimental and control groups, respectively. Symbols  $n_1$  and  $n_0$  denote the group sizes (so  $n_1 + n_0$  is the total sample size),  $\bar{y}_1$  and  $\bar{y}_0$  the sample means of the posttest scores, and  $s_1$  and  $s_0$  the sample standard deviations. The denominator in Equation 2 ( $s_p$ ) is the within-group pooled standard deviation of the posttest scores and estimates the population standard deviation in the groups, under the assumption of homogeneity of variance.<sup>1</sup>

Notably, Cohen’s  $d$  in Equation 2 represents the standardized mean difference between the control and the experimental group of the posttest in the sample as typically reported in randomized two-group postintervention test designs. It is important to realize this is just one member of a family of Cohen’s  $d$ s, and that reported effect size estimates are not always comparable (Lakens, 2013). In addition, this particular Cohen’s  $d$  is denoted in varying ways across sources such as books on meta-analysis and software packages (Lin & Aloe, 2020). Cohen (1988) was careful to differentiate between different effect sizes by

using subscripts. For instance, Cohen (1988) reported Cohen’s  $d$  in Equation 2 as  $d_s$ , where the  $s$  refers to sample. In this article, we will only focus on this particular Cohen’s  $d$  and drop the subscript  $s$ .

### Converting Cohen’s $d$ to the Point-Biserial Correlation

For MASEM, we need to convert Cohen’s  $d$  ( $d$ ) to a point-biserial correlation ( $r_{pb}$ ). Mathematically, the point-biserial correlation is no different than a Pearson product-moment correlation, but one that expresses the relationship between a natural dichotomous variable and a continuous variable (Lev, 1949; Tate, 1954). At the population level, the relationship between the standardized mean difference ( $\delta$ ) and the point-biserial correlation ( $\rho_{pb}$ ) can be expressed as follows (Bonett, 2020):

$$\rho_{pb} = \frac{\delta}{\sqrt{\delta^2 + \frac{1}{\pi(1-\pi)}}}. \quad (4)$$

And conversely:

$$\delta = \frac{\rho_{pb}}{\sqrt{(1 - \rho_{pb}^2)\pi(1 - \pi)}}. \quad (5)$$

<sup>1</sup> If primary studies do not report means and standard deviations, one can use the reported Cohen’s  $d$  if it is equal to the one in Equation 2. Sometimes means, standard deviations, or Cohen’s  $d$  is not reported but only the test statistics of the independent sample  $t$  test. Based on this information, one can also calculate Cohen’s  $d$ , see, for instance, Equation 2 in Lakens (2013) or Equation 2.5.3 in Cohen (1988).

In these expressions,  $\pi$  is the (hypothetical) proportion of individuals in the population assigned to the experimental condition, and  $1 - \pi$  is the proportion of individuals assigned to the control condition. Equation 4 shows that for a given  $\delta$ , the  $\rho_{pb}$  depends on  $\pi$ . A point-biserial correlation belonging to the (imaginary) situation in which the infinite population is randomly allocated to two groups, with an allocation ratio of 1:1 (effectively splitting the population into two infinite halves), implies that  $\pi = 1 - \pi = .50$  so that the expressions simplify to:

$$\rho_{pb} = \frac{\delta}{\sqrt{\delta^2 + 4}}, \quad (6)$$

and:

$$\delta = \frac{2\rho_{pb}}{\sqrt{1 - \rho_{pb}^2}}. \quad (7)$$

The relationship between Cohen's  $d$ —which we noted is a sample estimate ( $\hat{d}$ ) of the standardized mean difference in the population—and the sample point-biserial correlation ( $r_{pb}$ ) then becomes clear when we substitute the sample estimates in Equation 4:

$$\hat{\rho}_{pb} = \frac{\hat{\delta}}{\sqrt{\hat{\delta}^2 + \frac{1}{\hat{\pi}(1 - \hat{\pi})}}}, \quad (8)$$

and acknowledge (see formula 11 on p. 82 in Hedges & Olkin, 1985<sup>2</sup>) that:

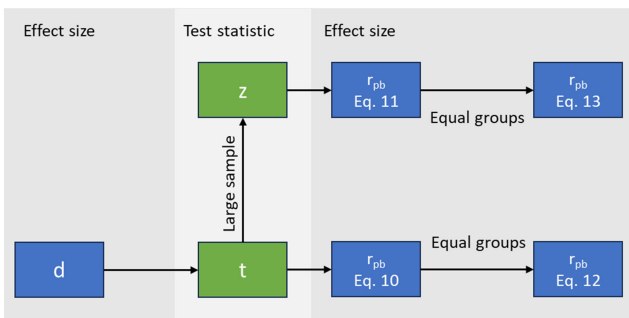
$$\hat{\delta} = d\sqrt{\frac{N - 2}{N}}. \quad (9)$$

Here,  $N$  is the total sample size, which is the sum of the two group sizes (which we denoted by  $n_0$  and  $n_1$ ). When you substitute Equation 9 into Equation 8, the relation can then be expressed as:

$$r_{pb} = \frac{d}{\sqrt{d^2 + \frac{(n_1 + n_0)^2 - 2(n_1 + n_0)}{n_1 n_0}}}. \quad (10)$$

The link between  $r_{pb}$  and  $d$  in a sample can also be explained through their significance tests (see also Aaron et al., 1998; Tate, 1954). Figure 1 shows the conceptual relations among the different

**Figure 1**  
Relationship Between Cohen's  $d$  and Point-Biserial Correlation ( $r_{pb}$ ) Through the  $t$ - and  $z$ -Test Statistics



Note. See the online article for the color version of this figure.

**Table 2**  
Overview of  $d$ -to- $r_{pb}$  Conversion Formulas

| $d$ -to- $r_{pb}$ conversion                       | General ( $t$ -based)                                                          | In the limit ( $z$ -based)                                            |
|----------------------------------------------------|--------------------------------------------------------------------------------|-----------------------------------------------------------------------|
| General                                            | $r_{pb} = \frac{d}{\sqrt{d^2 + \frac{(n_1 + n_0)^2 - 2(n_1 + n_0)}{n_1 n_0}}}$ | $r_{pb} \approx \frac{d}{\sqrt{d^2 + \frac{(n_1 + n_0)^2}{n_1 n_0}}}$ |
| Equally sized groups<br>$n_0 = n_1 = \frac{1}{2}N$ | $r_{pb} = \frac{d}{\sqrt{d^2 + 4 - \frac{8}{N}}}$                              | $r_{pb} \approx \frac{d}{\sqrt{d^2 + 4}}$                             |

Note.  $d$  = Cohen's  $d$ ,  $r_{pb}$  = point-biserial correlation,  $n_0$  = sample size of one group (e.g., control group) in a primary study,  $n_1$  = sample size of the other group (e.g., experimental group) in the same primary study,  $N = n_0 + n_1$  = total sample size in the primary study.

effect sizes and test statistics. The test statistics of both effect sizes follow a  $t$  distribution, which in large enough samples approaches the standard normal  $z$  distribution. Appendix A shows derivations through the test statistics yielding four expressions or approximations of the  $d$ -to- $r_{pb}$  conversion based on whether they lead to a point-biserial correlation for equal groups or not, and whether or not they are based on large sample approximations. Table 2 gives an overview of the four conversion formulas.

## General Conversion

Equation 10, which is also provided by, for example, Jacobs and Viechtbauer (2017) and Lakens (2013), is considered the most general conversion, as it does not assume a large sample size, and it applies to any ratio of group sizes. We refer to this equation as the “general conversion,” as the other three conversions are special cases of the general conversion.

## Large-Sample Conversion

As our derivations show, when the sample size is infinite or “large,” Equation 10 simplifies to:

$$r_{pb} \approx \frac{d}{\sqrt{d^2 + \frac{(n_1 + n_0)^2}{n_1 n_0}}}. \quad (11)$$

We will refer to this conversion as the “large-sample conversion.” This conversion has been suggested by, for example, Borenstein et al. (2009), Borenstein and Hedges (2019), and Lipsey and Wilson (2001).<sup>3</sup>

## Equal-Groups Conversions

With equal group sizes, the general conversion (Equation 10) simplifies to:

$$r_{pb} = \frac{d}{\sqrt{d^2 + 4 - \frac{8}{N}}}, \quad (12)$$

which is provided in Aaron et al. (1998), and we refer to this equation as the “equal-groups conversion.” The popular meta-analytic

<sup>2</sup> Note that the notation in Hedges and Olkin (1985) is different from how we provide it, but we changed it so that it is in line with the notation in our article.

<sup>3</sup> Note that these sources may differ in the notation of this formula.

handbooks have not discussed or recommended this equation, and, as far as we know, the commonly used statistical software or online conversion tools we considered have not implemented it (see Table 1).

### Large-Sample-Equal-Groups Conversion

A more well-known conversion is the large-sample version of Equation 12:

$$r_{pb} \approx \frac{d}{\sqrt{d^2 + 4}}. \quad (13)$$

This equation has also been recommended when group sizes are unknown (e.g., Borenstein & Hedges, 2019; Borenstein et al., 2009; Lipsey & Wilson, 2001). We refer to this equation as the “large-sample-equal-groups” conversion.

To summarize the above, we derived four expressions of the conversion from Cohen’s  $d$  to a point-biserial correlation. Three of those (Equations 10, 11, and 13) frequently appear in the meta-analytic literature, online tools, and statistical software. One is relatively unknown (Equation 12).

### Hedges’ $g$ : An Adjustment of Cohen’s $d$

Cohen’s  $d$  is known to yield biased estimates (overestimations) of the population mean difference, especially in small samples. Hedges’  $g$  ( $g$ ) is an often-used adjustment and can be approximated by (Hedges, 1981; Hedges & Olkin, 1985)<sup>4,5</sup>:

$$\text{Hedges' } g = \text{Cohen's } d \times \left(1 - \frac{3}{4 \times (n_0 + n_1) - 9}\right). \quad (14)$$

In practice, the difference between Hedges’  $g$  and Cohen’s  $d$  is small, especially when the sample size is large. Hedges’  $g$  is typically described as the corrected or unbiased version of Cohen’s  $d$  (e.g., Borenstein & Hedges, 2019; Borenstein et al., 2009; Hedges & Olkin, 1985), and additionally, it is sometimes advocated that it is preferable to report Hedges’  $g$  (e.g., Lakens, 2013). Yet, in meta-analyses, the aggregation of Hedges’  $g$  can yield (random effects) estimates that are stronger biased as compared to aggregation of Cohen’s  $d$ , especially when sample sizes are small (Hamman et al., 2018; Lin, 2018; Lin & Aloe, 2020). Therefore, in our study, we also considered whether the choice for Cohen’s  $d$  or Hedges’  $g$  matters for the estimation of the relevant parameters. We will evaluate the effect of substituting Hedges’  $g$  into the four conversion formulas instead of using Cohen’s  $d$ .

### Aim of the Simulation Study

In this study, we investigated the impact of the usage of the four  $d$ -to- $r_{pb}$  conversions and the adjustment to Hedges’  $g$  on the MASEM parameters using the R package metaSEM. In total, in line with the above, we considered eight methods to convert effect sizes from primary studies to point-biserial correlations. We evaluated the accuracy and precision of the MASEM parameters’ point estimates (i.e., pooled correlations, path coefficients, and the indirect effect) and the bias (i.e., lack of accuracy) in the standard errors of the path coefficients and indirect effect.

## Method

To investigate the effects of various conversion formulas on the pooled correlations, path coefficients, indirect effect, and the standard errors of the path coefficients and indirect effect, we conducted a Monte Carlo simulation study. We assumed the interest of a researcher would lie in a model in which a dichotomous variable  $X$  (intervention, with levels yes and no) has an effect on the continuous mediator  $M$  that in turn affects the continuous variable  $Y$ . We also allowed for a direct effect of  $X$  on  $Y$ . We investigated the quality of the parameter estimation under varying conditions, which comprised (a) the ratio of the sample sizes of the experimental and control group, (b) the number of primary studies included in the meta-analysis, (c) the sample size in these primary studies, and (d) the amount of missing correlations over studies. Below, we describe the different conditions in detail, as well as the data generation and evaluation criteria. All R scripts needed to reproduce the simulation can be found on the Open Science Framework as part of the additional online materials (<https://osf.io/ynkj3/>; de Jonge et al., 2024).

### Varying Conditions

#### Ratio

With respect to the ratio of the group sample sizes, we evaluated “exact,” “approximate,” and “mixed” conditions. In RCTs, the assignment to experimental or control group (allocation) is typically according to a ratio of 1:1, meaning researchers strive for a control group that is equal in size to the experimental group. The condition “exact” assumes this has been accomplished in all studies that are included in the meta-analysis. Due to circumstances that are beyond the control of the researcher, such as (differential) dropout, this may not always be accomplished in reality. In the “approximate” condition, we assumed group sizes are not exactly equal, although the allocation ratio was still 1:1. In this condition, we simulated that the group sizes were only approximately equal to each other (e.g., 62:58). In the “mixed condition,” we generated data according to a scenario in which half of the primary studies were assigned according to a 1:1 allocation ratio and in the other half to a 1:2 allocation ratio. It is not uncommon to find studies in which researchers aim for a 1:2 allocation ratio (e.g., Brethauer et al., 2022; Salmela-Aro et al., 2012). This can occur, for instance, when the intervention is expensive, but a sufficient number of participants needs to be included in order to have sufficient power to test for mean differences. The control group is then taken twice

<sup>4</sup> The adjustment of Cohen’s  $d$  is what we call Hedges’  $g$ . However, there is some confusion about this terminology (e.g., Lin & Aloe, 2020). Originally, Hedges (1981) and Hedges and Olkin (1985) called the expression in Equation 2  $g$  because it is based on the formula of Glass (1976). The introduced adjustment in Equation 14 was then called  $d$ . However, in later literature, letters were swapped (see, e.g., Borenstein et al., 2009, in which Larry Hedges is also a coauthor). Therefore, we refer to the expression in Equation 2 as Cohen’s  $d$  and the expression in Equation 14 as Hedges’  $g$ .

<sup>5</sup> While this expression is only an approximation of the original Hedges’  $g$ , it is the one that is typically reported in meta-analytic handbooks and literature (e.g., Borenstein & Hedges, 2019; Borenstein et al., 2009; Lakens, 2013) and used in substantive research as it is easier to apply than the original version. In addition, Lin and Aloe (2020) showed that when the total sample size is larger than 7, the original formula and approximation of this formula are almost equal to each other.

as large as the experimental group. In this mixed condition, we chose sample sizes always “approximately equal” according to their allocation ratio.

### Number of Studies

We varied the number of primary studies per meta-analysis as being 10, 30, or 50. These numbers fit both with the median number of primary studies included in meta-analysis as reported by Fernández-Castilla et al. (2020), and with numbers used in previous simulation studies (e.g., Lin & Aloe, 2020; Van Zundert & Miočević, 2020).

### Within-Study Sample Size

Sample size may differ across studies as well as across meta-analyses. We here varied the average sample size in a meta-analysis using averages of 26, 52, and 128. The number 128 reflects the required minimum sample size in many RCTs. This number is based on power calculations and refers to the situation in which a Cohen’s  $d$  of 0.50 (“a medium effect”<sup>6</sup> based on Cohen, 1988) can be detected with a power of least .80 (under the assumption the allocation ratio is 1:1). Similarly, the number 52 refers to minimal sample size that is required to detect a Cohen’s  $d$  of 0.80 (“a large effect” based on Cohen, 1988). Lastly, the number 26 represents the use of samples that can be considered small. We are interested in such samples because we aimed to gain knowledge about the behavior of the various conversion formulas, some of which assume sample sizes are large. Small sample sizes may violate this assumption. This way, we can investigate the robustness of the conversions. Furthermore, one strength of a meta-analysis is the increased power to detect effects by the aggregation of effects. Thus, it is not rare to find meta-analyses that include relatively small samples (e.g., Bijlsma et al., 2022; Leijten et al., 2019). For every primary study, we sampled the group sizes from uniform distributions, such that the total sample size equaled (or approximated) the average sample size in the given condition. The lower bound and upper bound were 20% lower and higher than the group size specified by the condition. These bounds represent a (maximum) dropout rate of 20% either taking or not taking into account this dropout in advance.

### Missing Correlations

In an ideal situation, each study provides information on all three variables ( $X$ ,  $M$ , and  $Y$ ). In practice, not all studies included will have investigated or reported on all three variables, so that missingness is common. Sheng et al. (2016) reported that, on average, 39% of the possible number of correlations in an MASEM is missing. We therefore imposed missing data by removing one of the variables ( $X$ ,  $M$ , or  $Y$ ) in a number of primary studies in such a way that the percentage of missing correlations equaled (or approximated) 39%.

The setup as described above yielded 54 different conditions; Group Sample Size Ratio (exact, approximate, mixed)  $\times$  Number of Primary Studies (10, 30, 50)  $\times$  Average Within-Study Sample Size (26, 52, 128)  $\times$  Percentage Missing Correlations (0%, 39%) =  $(3 \times 3 \times 3 \times 2 = 54)$ . In each condition, the number of replications was 2,000.

### Data Generation

In our simulation study, the primary studies represent RCTs in which the intervention affects two continuous variables ( $M$  and  $Y$ ). Assuming the intervention focuses on changing variable  $M$ , we specified a larger mean difference on variable  $M$  ( $\delta_M = .80$ ) than on variable  $Y$  ( $\delta_Y = .50$ ), representing respectively a large and medium effect according to the criteria of Cohen (1988). The values of .80 and .50 are also close to values of effect sizes that can be found in RCTs in the fields of psychiatry (e.g., Freeman et al., 2017; McGough & Faraone, 2009), child development (e.g., de Mooij et al., 2020; Leijten et al., 2019), psychological science (e.g., Berger et al., 2009), and medical science (e.g., Nissen et al., 2020; Small et al., 2018; Tikuišis et al., 2014; Wang et al., 2020). We assumed a within-group Pearson product-moment correlation between continuous variables  $M$  and  $Y$  of  $\rho_{YM} = .40$ . This value is close to Pearson product-moment correlation (.368) that requires a sample size of 52 for 80% power when performing a two-sided test of  $r = 0$ .

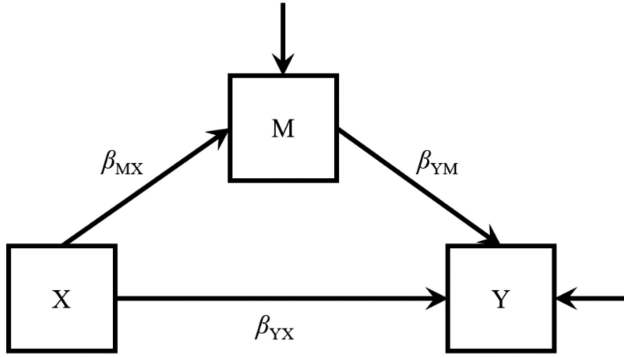
Based on these population values and using a value of  $\pi = .50$ , denoting equal group sizes in the population, we simulated data according to the following setup: means (and standard deviations) for  $M$  and  $Y$  were taken  $\mu_{M,1} = .40$  and  $\mu_{M,0} = -.40$  ( $\sigma_{M,1} = \sigma_{M,0} = 1$ ) and  $\mu_{Y,1} = .25$  and  $\mu_{Y,0} = -.25$  ( $\sigma_{Y,1} = \sigma_{Y,0} = 1$ )—which implies  $\delta_M = .80$  and  $\delta_Y = .50$ —and the within-group correlation between  $Y$  and  $M$ :  $r_{YM,1} = r_{YM,0} = .40$ . Here subscripts 1 and 0 refer to the experimental group and control group, respectively. As can be shown algebraically (see Appendix B), the population correlation coefficients then equal (rounded to two decimals)  $r_{pb, MX} = .37$ ,  $r_{pb, YX} = .24$ , and  $r_{YM} = .45$ , the standardized path coefficients  $\beta_{MX} = .37$ ,  $\beta_{YX} = .09$ , and  $\beta_{YM} = .42$ , and the indirect effect  $\beta_{MX} \times \beta_{YM} = .16$ . Against these coefficients, we compared the MASEM estimates.

For the between-study variance ( $\tau^2$ ) of both  $\delta_M$  and  $\delta_Y$ , we chose .10, which is close to respectively the mean and median between-study variance of standardized mean differences that Fernández-Castilla et al. (2020) reported for meta-analyses in behavioral and social science (.118 and .072), medical sciences (.109 and .120), and close to the values that Van Erp et al. (2017) reported for meta-analyses from psychological science (.099 and .040). We chose .02 for the between-study variance of  $r_{YM}$ , which is close to the mean and median for Pearson product-moment correlation coefficients as reported by Fernández-Castilla et al. (2020) for social science (.022 and .017, respectively) and Van Erp et al. (2017) for psychological science (.035 and .023, respectively).

We generated the data using R (Version 4.2.0; R Core Team, 2022) with the function `mvrnorm()` within the R package MASS (Venables & Ripley, 2002). We drew studies’ subpopulation parameter values from a multivariate normal distribution with mean vector ( $\delta_M = .80$ ,  $\delta_Y = .50$ ,  $r_{YM} = .40$ ) and associated between-study variances mentioned above. Based on the subpopulation’s Cohen’s  $d$ , we determined the means in the control and experimental groups and drew sample data from a multivariate normal distribution for every primary study separately for the experimental and control groups. Then we merged the data and,

<sup>6</sup> There is some critique on the use of the rules of thumb of Cohen (e.g., Funder & Ozer, 2019). However, the rules of thumb are still often used in practice.

**Figure 2**  
Path Model Representing Partial Mediation



depending on the condition, we deleted the number of variables that should be missing such that it led to (approximately) 39% missing correlations.

For each sample, we calculated the bivariate Pearson product-moment correlation between continuous variables  $M$  and  $Y$ . We also calculated standardized mean differences in  $M$  and  $Y$  using Equation 2 (Cohen's  $d$ ) and Equation 14 (Hedges'  $g$ ). We converted these  $d$ s and  $g$ s to point-biserial correlations using the four different conversion formulas (see Table 2). In this way, for each, we created four different sample correlation matrices based on Cohen's  $d$  and four different sample correlation matrices based on Hedges'  $g$ , leading to eight different meta-analytic data sets for each replication. To each meta-analytic data set, we fitted the partial mediation model depicted in Figure 2 with OSMASEM (Jak & Cheung, 2020) using the function `osmasem()` within the R package `metaSEM` (Cheung, 2015b). Perhaps to their surprise, the meta-analyst only has to provide correlation matrices and sample sizes and not, for example, the standard errors of the correlations or their sampling variance. In the background of `metaSEM` the variance and covariance matrix of the correlations are modeled as a function of the between-study and sampling (within-study) (co)variances. The default in `osmasem()` is the assumption that the between-study variance-covariance matrix,  $\mathbf{T}^2$ , is diagonal. Sampling variance and covariances are captured in  $\mathbf{V}_i$ . The sampling covariance is calculated using the formula of Olkin and Siotani (1976):

$$\begin{aligned} \text{cov}(\rho_{iAB}, \rho_{iCD}) = & (0.5\rho_{iAB}\rho_{iCD}(\rho_{iAC}^2 + \rho_{iAD}^2 \\ & + \rho_{iBC}^2 + \rho_{iBD}^2) + \rho_{iAC}\rho_{iBD} \\ & + \rho_{iAD}\rho_{iBC} - (\rho_{iAB}\rho_{iAC}\rho_{iAD} \\ & + \rho_{iAB}\rho_{iBC}\rho_{iBD} + \rho_{iAC}\rho_{iBC}\rho_{iCD} \\ & + \rho_{iAD}\rho_{iBD}\rho_{iCD}))/n_i. \end{aligned} \quad (15)$$

This expression refers to the population covariance between two correlation coefficients (in this case, between variables A and B and C and D). Therefore,  $\rho_i$  and  $n_i$  indicate the population correlation and the sample size in study  $i$ , respectively. The default in `metaSEM` is using the (weighted) mean correlations across samples as an estimate for each  $\rho_i$  (Furrow & Beretvas, 2005), such that eventually the sampling covariance matrices differ across studies only as a function of the study-specific sample sizes  $n_i$ .

The population variance of one correlation coefficient (e.g., between variables A and B) is the covariance with itself, for example,  $\text{cov}(\rho_{iAB}, \rho_{iAB})$ . The standard error of the correlations (either point-biserial correlations or Pearson product-moment correlations) is the square root of this variance. Note that this expression normally refers to the relation between two Pearson product-moment correlations (in this case, the correlation between variables A and B and the correlation between variables C and D) and does not necessarily apply to the point-biserial case. This forms one of the reasons to check the performance of `osmasem()` applied to RCT data.

## Evaluation Criteria

We evaluated the relative percentage bias (RB) in the pooled correlations, path coefficients, indirect effect, and the standard errors of the path coefficients and indirect effect, and the root-mean-square error (RMSE) of the estimated pooled correlations, path coefficients, and indirect effect. The following formula was used to quantify the RB in the pooled correlations:

$$\frac{\bar{r} - \rho}{\rho} \times 100\%. \quad (16)$$

In this formula,  $\bar{r}$  is the mean of the estimated pooled correlation across replications, and  $\rho$  is the true value of the pooled correlation.

The following formula was used to quantify the RB in the path coefficients:

$$\frac{\bar{\beta} - \beta}{\beta} \times 100\%. \quad (17)$$

In this formula,  $\bar{\beta}$  is the mean of the estimated path coefficients across replications, and  $\beta$  is the true value of the path coefficient. We considered RB values below 5% as acceptable (Hoogland & Boomsma, 1998). We estimated the RB in the indirect effect (i.e.,  $\beta_{MX} \times \beta_{YM}$ ) similarly. Since the indirect effect is the multiplication of two direct effects, we set the boundary of the indirect effect at 10.25% (because  $1.05 \times 1.05 = 1.1025$ ). We also calculated the Monte Carlo standard errors (MCSE) for the RB in all the pooled correlations, path coefficients, and indirect effect (Morris et al., 2019). This provides insight into whether the RB across the different conversion methods is meaningful or whether it is just noise. We used the following formula for the pooled correlations:

$$\text{MCSE} = \frac{SD\left(\frac{r_i - \bar{r}_i}{\bar{r}_i} \times 100\right)}{\sqrt{N_{\text{rep}}}}, \quad (18)$$

where  $SD$  is the standard deviation,  $N_{\text{rep}}$  represents the number of replications in our simulation study, and  $r_i$  the pooled correlation in one replication, and  $\bar{r}_i$  the mean of the pooled correlation across all replications. For the path coefficients and indirect effect, we replaced  $r_i$  and  $\bar{r}_i$  by the according parameter.

To estimate the RB in the standard errors of the path coefficient, we used:

$$\text{RB} = \frac{\overline{SE}(\hat{\beta}) - SD(\hat{\beta})}{SD(\hat{\beta})} \times 100\%, \quad (19)$$

where  $\overline{SE}(\hat{\beta})$  is the mean of the reported standard error of the

estimated path coefficients, and  $SD(\hat{\beta})$  the observed standard deviation of the estimated path coefficients across replications. To obtain a standard error of the indirect effect ( $X$  on  $Y$  through  $M$ ), we used Sobel's (1986) approximation (Kline, 2005):

$$SE_{ind} = \sqrt{\beta_{MX}^2 SE_{YM}^2 + \beta_{YM}^2 SE_{MX}^2}, \quad (20)$$

in which  $SE_{YM}$  and  $SE_{MX}$  are the standard errors of the estimated  $\beta_{YM}$  and  $\beta_{MX}$ , respectively. RB in the standard error of the indirect effect was calculated using the standard deviation of the product of estimated path coefficients across replications:  $SD(\hat{\beta}_{MX}\hat{\beta}_{YM})$ . We considered bias in the standard errors below 10% acceptable (Hoogland & Boomsma, 1998). These criteria of Hoogland and Boomsma (1998) are typically applied in simulation studies on meta-analysis (e.g., Cai & Fan, 2020; Gnambis & Sengewald, 2023; Moeyaert et al., 2017).

We complement the RB with the RMSE, which is a measure of overall estimation quality, as it accounts for both bias and variability around the population value. The formula to calculate the RMSE is:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N_{rep}} (\hat{\beta}_i - \beta)^2}{N_{rep}}}. \quad (21)$$

Mathematically, this is equal to  $RMSE = \sqrt{(\bar{\hat{\beta}} - \beta)^2 + VAR(\hat{\beta})}$ , where  $(\bar{\hat{\beta}} - \beta)$  represents the absolute bias, and  $VAR(\hat{\beta})$  the variance in estimates of  $\beta$ . This clarifies that the RMSE combines information on bias (systematic deviations from the true value) and precision (spread or variability), with lower RMSE values representing

better overall performance. However, an unbiased estimator can still have a high RMSE if its variance is large, meaning that the accuracy in terms of RMSE requires both low bias and low variance. If an estimator is unbiased, the RMSE equals the standard deviation of the estimates (i.e., empirical standard error). In a similar vein, we estimated the RMSEs in the indirect effect and pooled correlations.

## Results

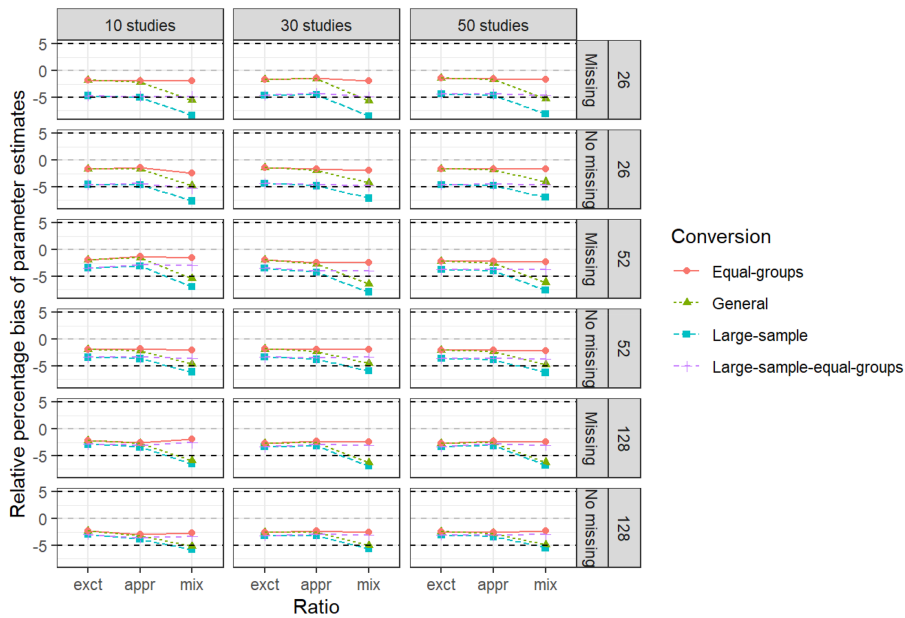
### Preliminary Checks: Nonconvergence and Inadmissible Solutions

Prior to the statistical analysis, we performed two quality checks. As a first check, we obtained the used software's (i.e., R package metaSEM) convergence status and the parameter point estimates (correlations and path coefficients) and accompanying standard errors. According to this information, nonconvergence was absent. In addition, parameter estimates were never implausible (e.g., a correlation larger than 1). Yet, in some replications, the reported standard errors were extremely large (e.g., larger than 100). If this occurs in practice, a change of starting values and rerunning the model typically solves the issue. To ensure that such extreme cases did not influence the results of our simulation study, we used the median standard error instead of the mean standard error in Equation 19 to evaluate the bias in the standard errors.

The second quality check consisted of comparing the model-implied correlation matrix with the pooled correlation matrix. We obtained the latter from the first stage of two-stage SEM

**Figure 3**

*Relative Percentage Bias in  $r_{pb,MX}$  in the 54 Conditions When Converting Cohen's  $d$  With the Four Conversions*



*Note.* The numbers 26, 52, and 128 on the right side denote the within-study sample sizes. The dotted black lines indicate the margins of acceptable values of relative percentage bias. Abbreviations exct, appr, and mix on the x axis represent exact, approximate, and mixed condition, respectively. See the online article for the color version of this figure.

(Cheung, 2014), using the function `tssem1()` and its default settings within the R package `metaSEM`. Since the fitted path model is saturated, the model-implied correlation matrix should be equal to this pooled correlation matrix. Any difference between the two matrices served as an indicator of an issue with the optimization procedure. Whenever any element in the model implied correlation matrix differed more than .005 from the corresponding element in the pooled correlation matrix, we judged the solution as inadmissible and omitted the replication from further statistical analysis. This criterion resulted in the exclusion of four replications.

## RB in Pooled Correlations

In summary, RB in the pooled correlation estimates were typically negative and averaged over conditions appeared smaller than 5% (see S1 in the online supplemental materials for tables with full RB results, including the MCSE). However, clear patterns of presence and absence of bias emerged when we considered the conditions separately. These patterns depended to a certain extent on which parameter we considered.

For  $r_{pb,MX}$  and  $r_{pb,YX}$ , the results were quite similar. The RB was consistently closer to zero when based on Cohen's  $d$  rather than Hedges'  $g$  but showed the same patterns across conditions. We focus on summarizing the results for Cohen's  $d$ . Figures 3 and 4 show RB in  $r_{pb,MX}$  and  $r_{pb,YX}$  obtained after the conversion from  $d$  to  $r_{pb}$  using the four different conversion formulas in each of the 54 conditions. In these figures, we see that the use of the equal-groups conversion typically yielded the most accurate estimates. This conversion was also the only conversion that never led to

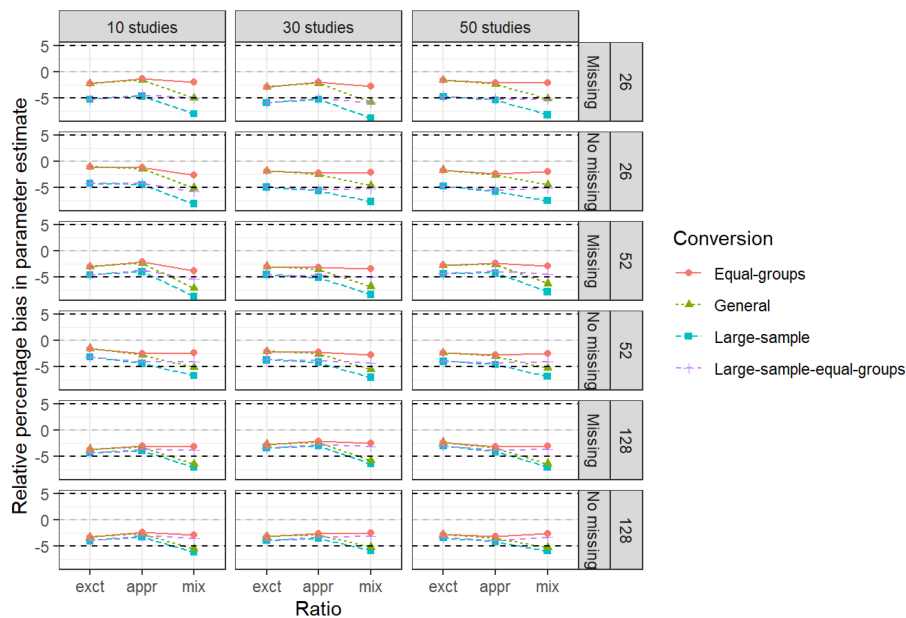
bias in any of the conditions. We also see that the four conversions yielded similar results in the exact and approximate conditions but not in the mixed conditions. In the (18) exact and (18) approximate conditions, we found the following. On average across those  $(18 + 18 = 36)$  conditions, we found the lowest RB when using the equal-groups conversion. The general conversion also performed well, whereas the large-sample-equal-groups conversion and the large-sample conversion performed somewhat worse. On average across the 36 conditions, the RB always remained within 5%. However, for the latter two conversion, we found bias in some conditions. Especially, in  $r_{pb,YX}$ , when the sample size was small.

However, in the (18) mixed conditions, as Figures 3 and 4 show, this did not hold all true. The effects of the different conversions were more pronounced. On average, the equal-groups conversion still led to the lowest RBs (average RB in  $r_{pb,MX} = -2.11\%$  and average RB in  $r_{pb,YX} = -2.66\%$ ), remaining within the 5%. The large-sample-equal-groups conversion performed well on average across the 18 conditions (average RB in  $r_{pb,MX} = -3.80$  and average RB in  $r_{pb,YX} = -4.46$ ). Yet, with the latter conversion, bias appeared in some conditions, in particular in  $r_{pb,YX}$ , and when the sample size was small. The general conversion clearly led to biased results (on average across 18 conditions, the RB in  $r_{pb,MX}$  equaled  $-5.24\%$  and in  $r_{pb,YX}$   $-5.61\%$ ) and so did the large-sample conversion (on average, RB in  $r_{pb,MX}$  equaled  $-6.90\%$  and RB in  $r_{pb,YX}$   $-7.37\%$ ).

In conclusion, the effects of the conversions were most pronounced in the mixed sample conditions. These differences are illustrated by the distribution plots of  $r_{pb,MX}$  and  $r_{pb,YX}$  in Figures 5 and 6. These figures also clearly demonstrate that estimates based on the conversion of effect size Cohen's  $d$  were (on average) more accurate

**Figure 4**

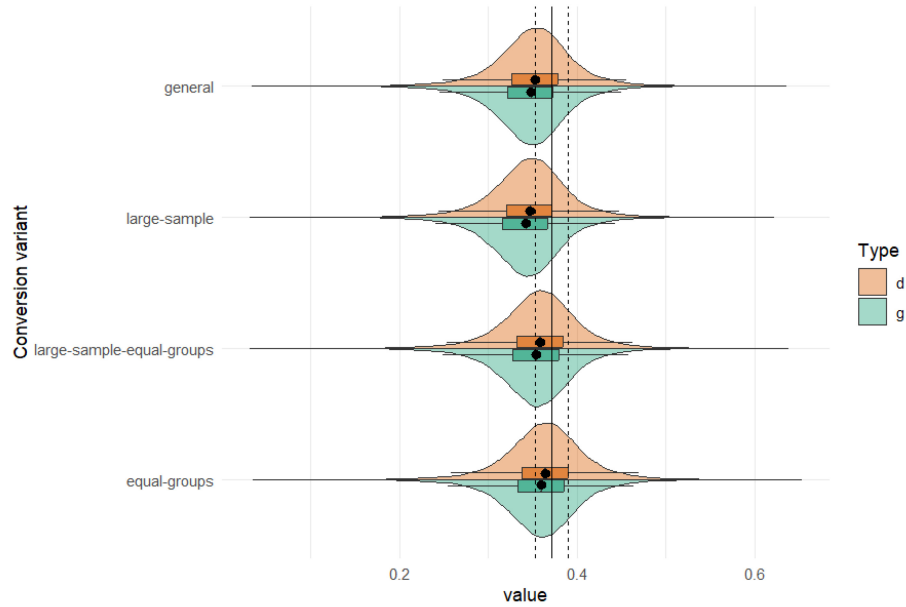
*Relative Percentage Bias in  $r_{pb,YX}$  in the 54 Conditions When Converting Cohen's  $d$  With the Four Conversions*



*Note.* The numbers 26, 52, and 128 on the right side denote the within-study sample sizes. The dotted black lines indicate the margins of acceptable values of relative percentage bias. Abbreviations *exct*, *appr*, and *mix* on the  $x$  axis represent exact, approximate, and mixed condition, respectively. See the online article for the color version of this figure.

**Figure 5**

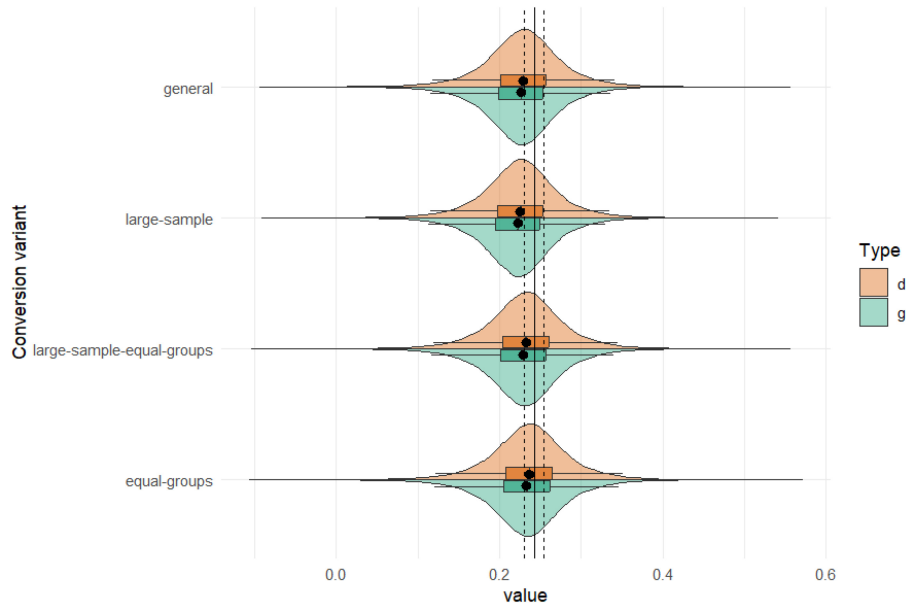
*Distributions of  $r_{pb,MX}$  as Obtained With the Four Conversions From Cohen's  $d$  (Orange, Top) or Hedges'  $g$  (Green, Bottom Upside Down) in All Replications of the 18 Mixed Conditions*



*Note.* The straight vertical line shows the true value of  $r_{pb,MX}$ , and the two dashed vertical lines show the margins of acceptable bias. So, if the mean of the distribution (black dot) falls within these margins, we consider the estimation unbiased. See the online article for the color version of this figure.

**Figure 6**

*Distributions of  $r_{pb,YX}$  as Obtained With the Four Conversions From Cohen's  $d$  (Orange, Top) or Hedges'  $g$  (Green, Bottom Upside Down) in All Replications of the 18 Mixed Conditions*



*Note.* The straight vertical line shows the true value of  $r_{pb,YX}$ , and the two dashed vertical lines show the margins of acceptable bias. So, if the mean of the distribution (black dot) falls within these margins, we consider the estimation unbiased. See the online article for the color version of this figure.

than estimates based on the conversion of effect size Hedges'  $g$ , while the equal-groups conversion led to the most accurate estimates. Also visible is that the accuracy depended on the conversion, yet the dispersion, hence the precision, did not. As expected, the choice for Cohen's  $d$  or Hedges'  $g$  or the type of conversion used for the other relations had no effect on  $r_{YM}$ . We never found bias in any of the conditions (i.e., RB values always below the boundary of 5%). For comparison, the overall average RB was  $-3.19\%$ .

### RB in Path Coefficients and Indirect Effect

Having gained knowledge about bias in the correlation coefficients, we can turn to the bias in the path coefficient and the indirect effect (see S2 in the online supplemental materials for tables with full RB results, including the MCSE). The path coefficients in our (saturated) model are functions of the correlations. Specifically,  $\beta_{MX} = \rho_{pb,MX}$ ,  $\beta_{YM} = (\rho_{pb,MX}\rho_{pb,YX} - \rho_{YM})/(\rho_{pb,MX}^2 - 1)$ , and  $\beta_{YX} = (\rho_{pb,MX}\rho_{YM} - \rho_{pb,YX})/(\rho_{pb,MX}^2 - 1)$ . Consequently, the results for  $\hat{\beta}_{MX}$ , the direct effect of  $X$  on  $M$ , are the same as for  $r_{pb,MX}$ , and we refer to the above. The results for the other path coefficients and the indirect effect are discussed below.

As Figure 7 shows, RB values of  $\hat{\beta}_{YX}$  did not show such a clear pattern as in any of the other parameter estimates, which complicates the interpretation. We attribute this to the fact that  $\beta_{YX}$  is a complex function of the parameters  $\rho_{pb,MX}$ ,  $\rho_{pb,YX}$ , and  $\rho_{YM}$ , such that bias in  $\beta_{YX}$  is a complex function of bias in  $\rho_{pb,MX}$  and bias in  $\rho_{pb,YX}$  (and bias in  $\rho_{YM}$ ). Overall, we can draw the following conclusions. While RB values in all other parameter estimates are negative, denoting a (tendency) of systematic underestimation of the population

value, RB values of  $\hat{\beta}_{YX}$  could also turn out positive. Hence, RB values of  $\hat{\beta}_{YX}$  showed both under- and overestimation, depending on the condition. We did not find the same pattern as for the other parameter estimates. First, on average across the conditions, Cohen's  $d$  did not always outperform Hedges'  $g$ . Second, the equal-groups conversion did not generally outperform the other conversion formulas.

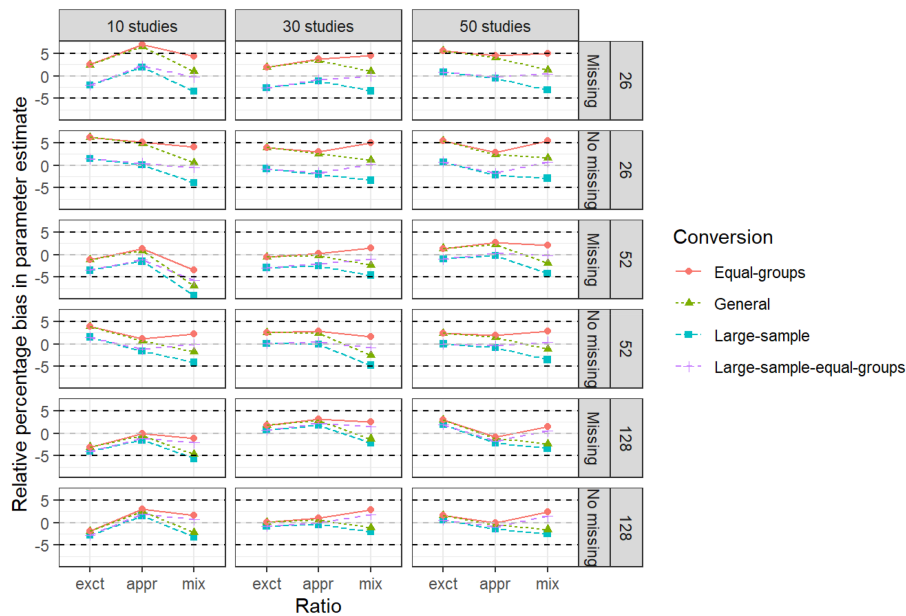
However, as the distribution plot in Figure 8 illustrates, even in the mixed-sample conditions (where the differences between the conversions are generally the largest), the distributions of  $\hat{\beta}_{YX}$  for the different conversions hardly differed from each other. We also see that the RB values of  $\hat{\beta}_{YX}$  in the mixed conditions typically did not exceed the boundary of 5%. Hence, the use of different conversions had limited effect on the estimation of  $\beta_{YX}$ . The conversions also had little effect on the estimation of  $\beta_{YM}$  as Figures 9 and 10 illustrate. We did not find bias in  $\hat{\beta}_{YM}$  (except for one condition with a coincidental RB value of 5.13%), and the distributions are approximately equal, no matter which conversion or effect size (Cohen's  $d$  or Hedges'  $g$ ) was used.

The indirect effect is the product of  $\beta_{MX}$  and  $\beta_{YM}$ . As the different conversions did not affect the estimation of  $\beta_{YM}$ , the effects of the conversions on the indirect effect resembled the effects of the conversion on the estimation of  $\beta_{MX}$ . Again, converting Cohen's  $d$  using the equal-groups conversion led on average to the smallest RB, while in the mixed conditions, we found the most pronounced difference (see Figure 11).

### RB in Standard Errors

As the interest lied in path modeling, not only bias in the point estimates of the parameters in the path is relevant, but also bias in

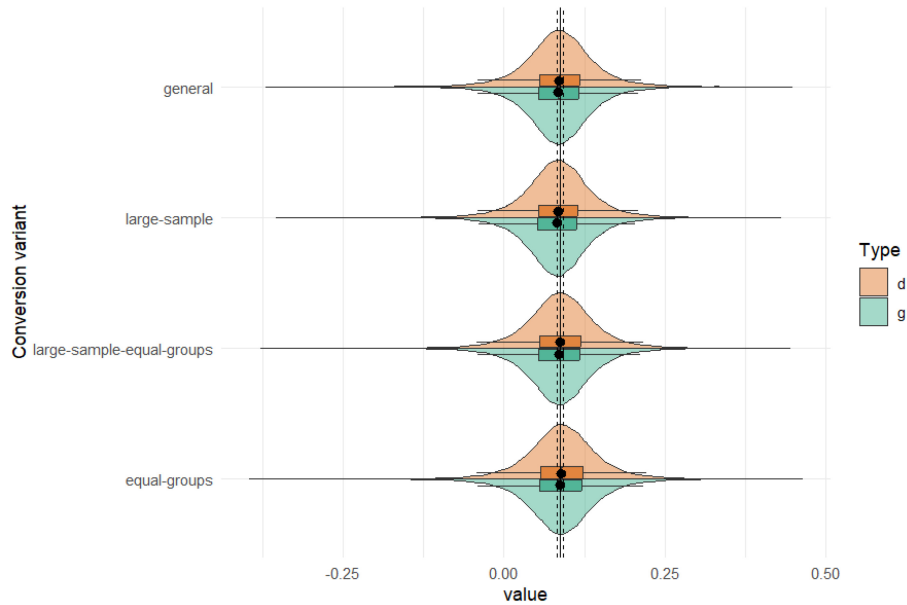
**Figure 7**  
*Relative Percentage Bias in  $\hat{\beta}_{YX}$  in the 54 Conditions When Converting Cohen's  $d$  With the Four Conversions*



*Note.* The numbers 26, 52, and 128 on the right side denote the within-study sample sizes. The dotted black lines indicate the margins of acceptable values of relative percentage bias. Abbreviations exact, appr, and mix on the  $x$  axis represent exact, approximate, and mixed condition, respectively. See the online article for the color version of this figure.

**Figure 8**

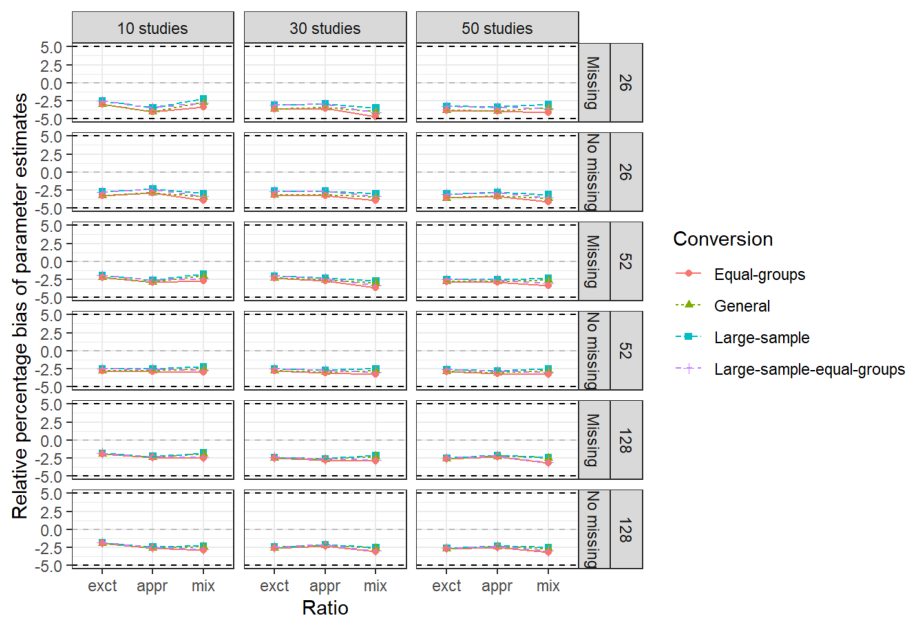
Distributions of  $\hat{\beta}_{YX}$  as Obtained With the Four Conversions From Cohen's  $d$  (Orange, Top) or Hedges'  $g$  (Green, Bottom Upside Down) in All Replications of the 18 Mixed Conditions



*Note.* The straight vertical line shows the true value of  $\beta_{YX}$ , and the two dashed vertical lines show the margins of acceptable bias. So, if the mean of the distribution (black dot) falls within these margins, we consider the estimation unbiased. See the online article for the color version of this figure.

**Figure 9**

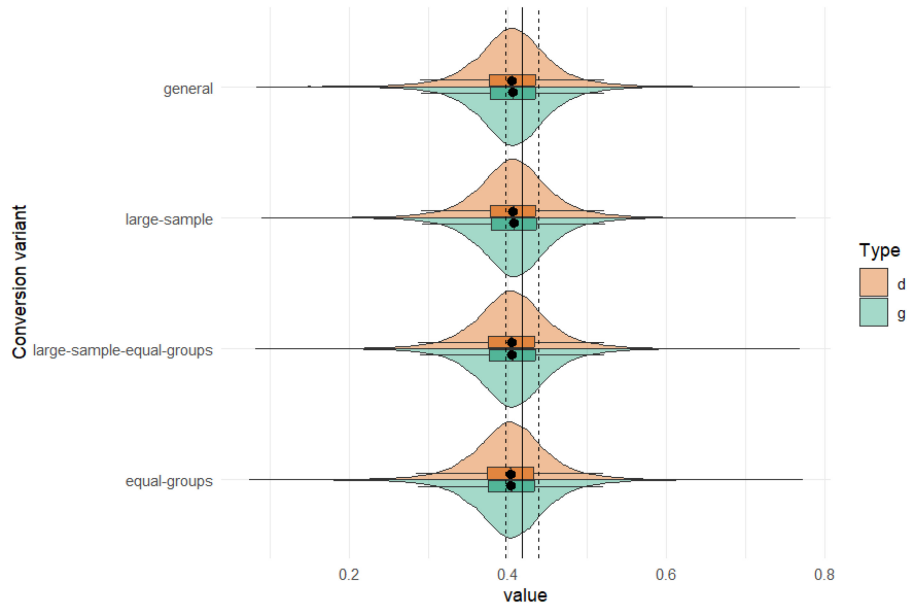
Relative Percentage Bias in  $\hat{\beta}_{YM}$  in the 54 Conditions When Converting Cohen's  $d$  With the Four Conversions



*Note.* The numbers 26, 52, and 128 on the right side denote the within-study sample sizes. The dotted black lines indicate the margins of acceptable values of relative percentage bias. Abbreviations exct, appr, and mix on the  $x$  axis represent exact, approximate, and mixed condition, respectively. See the online article for the color version of this figure.

**Figure 10**

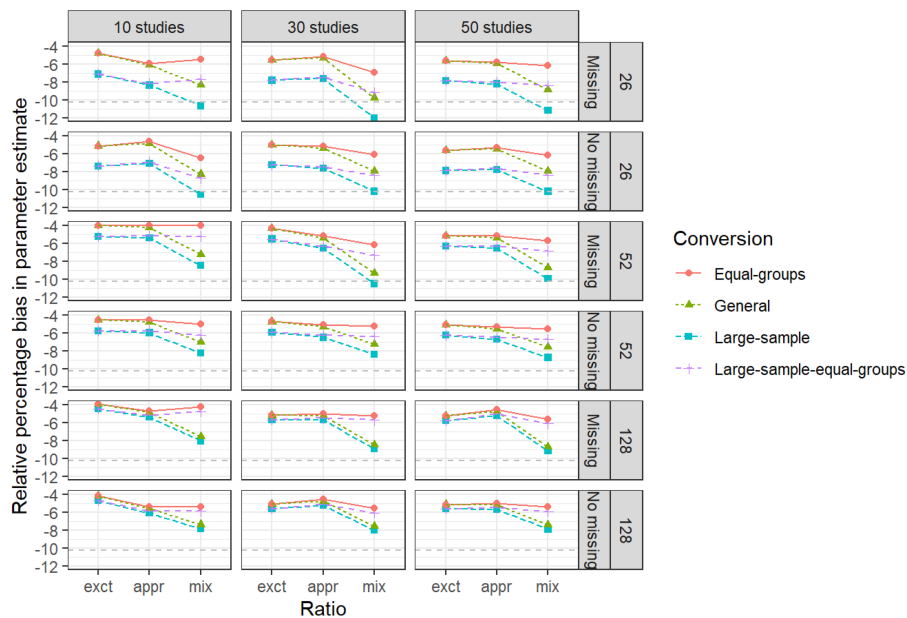
Distributions of  $\hat{\beta}_{YM}$  as Obtained With the Four Conversions From Cohen's  $d$  (Orange, Top) or Hedges'  $g$  (Green, Bottom Upside Down) in All Replications of the 18 Mixed Conditions



*Note.* The straight vertical line shows the true value of  $\beta_{YM}$ , and the two dashed vertical lines show the margins of acceptable bias. So, if the mean of the distribution (black dot) falls within these margins, we consider the estimation unbiased. See the online article for the color version of this figure.

**Figure 11**

Relative Percentage Bias in Indirect Effect in the 54 Conditions When Converting Cohen's  $d$  With the Four Conversions



*Note.* The numbers 26, 52, and 128 on the right side denote the within-study sample sizes. The dotted black lines indicate the margins of acceptable values of relative percentage bias. Abbreviations exct, appr, and mix on the  $x$  axis represent exact, approximate, and mixed condition, respectively. See the online article for the color version of this figure.

their standard errors (see S3 in the online supplemental materials for tables with full RB results). The RB values of the standard errors of the path coefficients and indirect effect were often negative, denoting a tendency of underestimation. Nevertheless, RB values most often remained within the 10% limit, irrespective of whether Cohen's  $d$  or Hedges'  $g$  was used, and no matter which conversion was applied. There were no apparent differences across conversions, regardless of whether Cohen's  $d$  or Hedges'  $g$  was plugged in (for figures, see S4 in the online supplemental materials). For  $\hat{\beta}_{YX}$  and  $\hat{\beta}_{YM}$ , the RB values for the standard errors were always below 10%. For  $\hat{\beta}_{MX}$  and the estimates of the indirect effect, in a few conditions we found bias, or RB values slightly more extreme than the boundary of 10%, for all conversions (either  $d$ -based or  $g$ -based). These conditions all concerned the inclusion of a small number of 10 primary studies in the meta-analysis.

## RMSE

We also considered the RMSE for the pooled correlations, path coefficients, and indirect effect. As expected, overall the RMSE was typically larger in conditions with smaller number of studies, smaller sample sizes, and when data were missing. Differences between the four conversions were relatively small, but the RMSE generally showed the same pattern as reported for the RB in the parameter estimates. Figure 12 provides an illustration and shows the RMSE of the

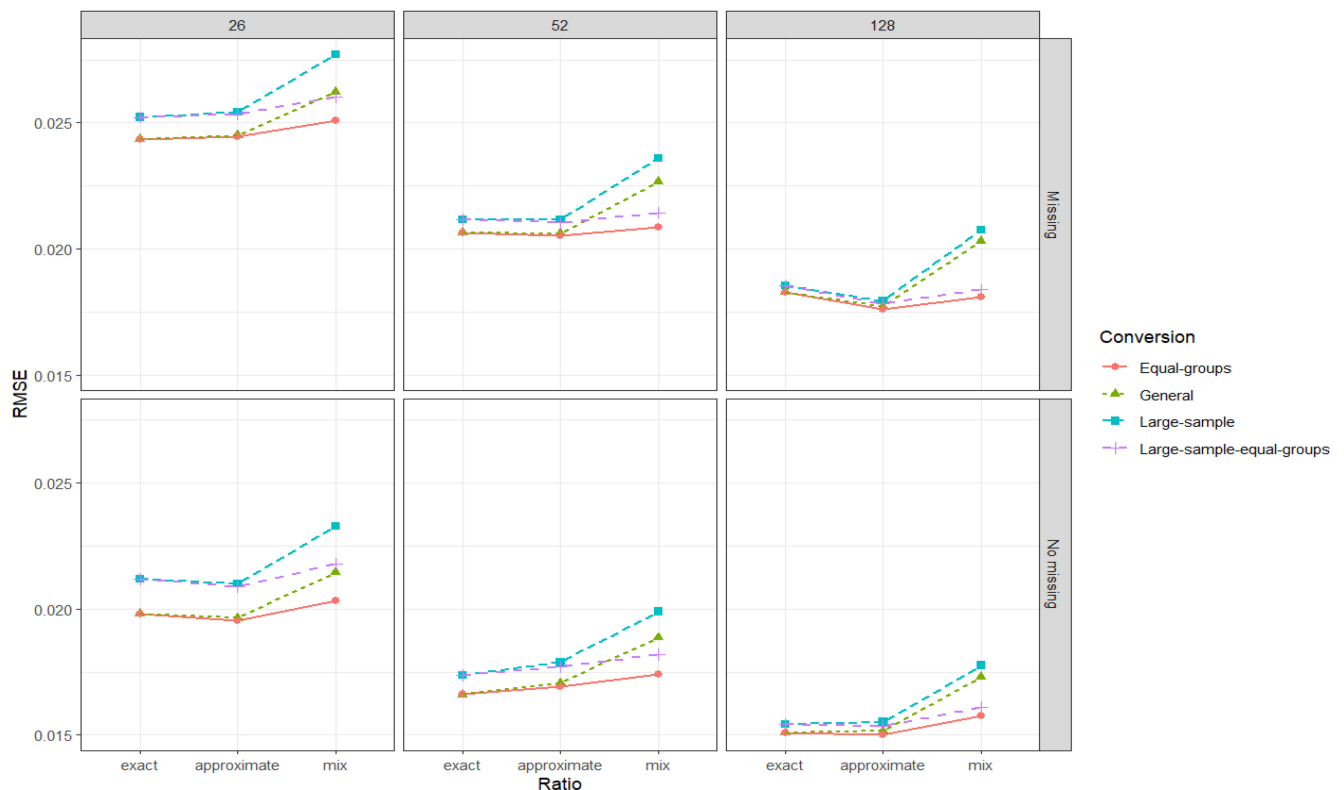
indirect effect when converting Cohen's  $d$  in conditions with 50 studies. We see that overall, the equal-groups conversion performs best. In conditions with a sample size of 26, we see that the large-sample conversion and the large-sample-equal-groups conversion typically yielded the largest RMSE values. In the mixed conditions, we found the strongest differences between the conversions. The results for all parameters in all conditions can be found in S5 and S6 in the online supplemental materials.

## Discussion

We conducted a simulation study to investigate the effect of using different conversion formulas on the estimation of parameters in MASEM of RCTs. Based on the results of this simulation study, we advise the use of the equal-groups conversion (Equation 12) to convert Cohen's  $d$  to a point-biserial correlation when the interest lies in MASEM analysis with RCTs. Overall, this conversion typically resulted in the most accurate estimates of the pooled correlation, path coefficients, and indirect effect. Precision did not depend on the type of conversion. The accuracy of the standard errors of the path coefficients and indirect effect did not depend on the conversion either. However, in line with previous research, we found that when the number of primary studies is small, the standard errors can be biased (e.g., Jak & Cheung, 2020), but this does not depend on which  $d$ -to- $r_{pb}$  conversion is applied.

**Figure 12**

*RMSE of the Indirect Effect as Obtained When Converting Cohen's  $d$  With the Four Conversions in Conditions With 50 Primary Studies*



*Note.* The numbers 26, 52, and 128 on the top denote the within-study sample size conditions. RMSE = root-mean-square error. See the online article for the color version of this figure.

We seem to be the first to advise the use of this equal-groups conversion. In prior meta-analytic literature, the general conversion (e.g., Jacobs & Viechtbauer, 2017; Lakens, 2013) or the large-sample conversions have been recommended (e.g., Borenstein & Hedges, 2019; Borenstein et al., 2009; Lipsey & Wilson, 2001). The reason for the discrepancy in advice is that we here specifically focus on RCT data. The equal-groups conversion works best for this application because this conversion “corrects” for the inference to the (imaginary) situation in which the infinite population is randomly allocated to an experimental and control group with an allocation ratio of 1:1 (effectively splitting the population into two infinite halves). That is, the equal-groups conversion will convert Cohen’s  $d$  to that particular point-biserial correlation that would have been expected when the parameter  $\pi$  (proportion between the groups in the population) in Equation 8 is .50. To convert Cohen’s  $d$  to a point-biserial correlation using the equal-groups conversion, information about the total sample size from the primary studies is required. In contrast, the large-sample-equal-groups conversion does not require this information, yet still converts to the situation where  $\pi$  is .50. This conversion generally performs worse than the equal-groups conversion in small-sample contexts and may even introduce bias. However, under our conditions, we found that when the average sample size across all primary studies is (at least) 128, the large-sample-equal-groups conversion yields unbiased pooled correlations and path coefficients. In those situations, one might choose to use the large-sample-equal-groups conversion instead of the equal-groups conversion. Further research is needed to determine under which other conditions the large-sample-equal-groups conversion yields unbiased MASEM estimates.

Note that in general, a point-biserial correlation depends on parameter  $\pi$ , while a standardized mean difference does not. This means that in order to make the inference to the population point-biserial correlation, a sample point-biserial correlation should not be calculated based on the sample distributions but on the so-called (assumed) population “base rate” (i.e., the prevalence; McGrath & Meyer, 2006). In an RCT, the natural target base rate is .50 because of the rationale behind RCT, as explained in the introductory part of the article. This implies that the group sample sizes should either reflect the population “base rate,” meaning that for RCTs, the sample sizes within the primary studies included in the meta-analysis should be exactly equal to each other, or a correction should be applied. If the group sizes are unequal, one can “correct” for this unbalance by converting Cohen’s  $d$  to a point-biserial correlation with the equal-groups conversion. If such a correction is not applied, as happens when using the general or large-sample conversion, this implies different values of  $\pi$  are used for different primary studies. A meta-analysis of these point-biserial correlations would yield an uninterpretable pooled point-biserial correlation, because there is no specific  $\pi$  attached. This is why McGrath and Meyer (2006) stress that in a meta-analysis of point-biserial correlations, the  $d$ -to- $r_{pb}$  conversion should be done on the basis of a specific “target base rate.”

We also note that when  $\pi$  is (assumed) to be .50, the point-biserial correlation is at its maximum (absolute) value. When the groups in the sample are unequal and one would convert Cohen’s  $d$  using the general or large-sample conversion, the outcome would be an underestimation of the strength of the point-biserial correlation in the population. It is important to note that the distributions over groups in a sample can differ substantially from the 1:1

hypothesized distribution in the population when applying an RCT. The extremer this difference, the smaller the corresponding point-biserial correlation (McGrath & Meyer, 2006). This is why in the mixed conditions in our simulation study—when the allocation ratio was 1:2 in half of the number of studies—we found the differences between the conversions the most pronounced. When the groups are exactly equal to each other, the general conversion and equal-groups conversion will yield the same point-biserial correlation and no correction is needed.

## Caution

Although our general recommendation is to convert Cohen’s  $d$  to a point-biserial correlation using the equal-groups conversion for RCT data, this conversion does not guarantee that path coefficients will remain unbiased under more extreme conditions (e.g., very high or low values of the standardized mean difference) than those examined in our simulation study. Because the conversion function from  $d$  to  $r_{pb}$  is monotonic, the pattern of pooled correlations remains consistent, and the equal-groups conversion generally provides the most accurate estimate at that level. However, it is important to note that path coefficients and indirect effects may become biased under other conditions than investigated in this simulation study. In some cases, this bias might, coincidentally, be even greater when using the equal-groups conversion than when using alternative conversion methods. This arises from the fact that path coefficients are nonlinear and nonmonotonic functions of the pooled correlations, which means that even small differences in the input correlations can lead to complex or disproportionate changes in the resulting estimates.

## Hedges’ $g$ and Cohen’s $d$

Our simulation showed that using Hedges’  $g$  generally resulted in larger RB in the point estimates than using Cohen’s  $d$ . This result aligns with other recent simulation studies that show that in random effects meta-analysis the aggregation of Hedges’  $g$  can lead to larger bias than the aggregation of Cohen’s  $d$ , in particular when sample sizes are small (Hamman et al., 2018; Lin, 2018; Lin & Aloe, 2020). Such larger bias may arise from multiple sources. First, the way the variance is estimated, and in turn how studies are weighted to pool the correlations, can be of influence (Lin, 2018; Lin & Aloe, 2020). Second, the results of our simulation suggest that the conversion from a standardized mean difference to the point-biserial correlation plays a role. This conversion is nonlinear, resulting in underestimation of the point-biserial correlation. Since Hedges’  $g$  is developed to correct for overestimation, the effect after conversion is a further underestimation of pooled correlations. Yet, this does not necessarily lead to underestimation of path coefficients: Since path coefficients are complex functions of the pooled correlations, the bias in the path coefficient is a complex function of the bias in the pooled correlations, while the size of the bias in the pooled correlation will depend on the size of the standardized mean difference (i.e., the smaller the standardized mean difference, the smaller the underestimation). Future simulation studies may help to provide more insight as to why and under what conditions the aggregation of Hedges’  $g$  is less accurate than the aggregation of Cohen’s  $d$  when considering the pooled correlations and path coefficients in an MASEM.

## Correcting the Point-Biserial Correlation for Unequal Groups

In our study, we considered the situation in which primary studies report means, standard deviations, or standardized mean differences (i.e., Cohen's  $d$  and Hedges'  $g$ ). This is common practice, since it is challenging to access raw data (e.g., Gabelica et al., 2022; Wicherts et al., 2006). Instead of the aforementioned summary statistics, point-biserial correlations between the intervention variable and continuous variables may have been reported in primary studies. Before including such coefficients in MASEM, one would need to check if the group sizes of the experimental and control groups were exactly equal. If not, one can correct the reported point-biserial correlation by first converting it to Cohen's  $d$ , using the inverse of the general conversion:

$$d = \frac{\sqrt{\frac{(n_1 + n_0)^2 - (2(n_1 + n_0))}{(n_1 n_0)}} \times r_{pb}}{\sqrt{1 - r_{pb}^2}}. \quad (22)$$

This Cohen's  $d$  can then be converted back into a point-biserial correlation using the equal-groups conversion. Note that this procedure differs from the correction that Hunter and Schmidt (2004, p. 280) provide to correct the point-biserial correlation for unequal group sample sizes to a point-biserial correlation that one would have obtained when the group sample sizes would have been equal to each other.

## Natural Dichotomous or Artificially Dichotomized Predictor Variables

In our study, we focused on RCT data and on the situation in which the predictor variable reflects whether participants are allocated to an experimental or control group. We stress that our advice to use the equal-groups conversion does not (necessarily) apply to designs without randomization. We may think of situations in which the dichotomous variables represent having a college degree or not, being pregnant or not, having children or not, having a specific disease or not, or school dropout or not. These can be considered natural dichotomies, but in such instances, the distribution in the population is unlikely to be according to a 1:1 ratio (i.e., prevalence of .50). The equal-groups conversion is then not applicable, because the statistical inference should not be toward .50, but to the (known or estimated) prevalence in the population. This means that when converting Cohen's  $d$  to a point-biserial correlation in a meta-analysis, one should apply a conversion that takes into account this prevalence. To investigate the effect of including natural dichotomous variables within MASEM, we recommend further research. Besides natural dichotomous variables, one could also be interested in nonnatural dichotomous variables (i.e., artificially dichotomized variables). A dichotomized variable represents, for example, whether an individual scored above or below a cutoff score on a continuous variable, such as a sum score on an anxiety or depression questionnaire. For the inclusion of artificially dichotomized variables in MASEM, one would need an additional conversion of the point-biserial to the biserial correlation (de Jonge et al., 2020).

## Effect Size Calculator and Converter (ESCACO)

We recommend the equal-groups conversion, but—as far as we know—no statistical software or online conversion tool would assist in such conversion. To assist researchers with the conversion of effect sizes suitable for application in MASEM, we present a user-friendly web application entitled ESCACO. This “Shiny app” can be found at <https://hdejonge.shinyapps.io/ESCACO> and the underlying R script on the Open Science Framework as part of the additional online materials at <https://osf.io/ynkj3/>, so that one can also use the ESCACO offline. Users can enter Cohen's  $d$  and sample size from a primary study, and the application will return the  $d$ -to- $r_{pb}$  converted effect size based on the equal-groups conversion. If a primary study does not provide Cohen's  $d$ , the application also provides a section to get the needed Cohen's  $d$  based on means and standard deviations or based on the  $t$  value, using the formulas described in the introductory part of the article. If a primary study provides Hedges'  $g$ , the application has a subsection to convert it back to Cohen's  $d$ . The formula that we implemented to do so is as follows:

$$\text{Cohen's } d = \frac{\text{Hedges' } g}{1 - \left( \frac{3}{4 \times (n_1 + n_0) - 9} \right)}. \quad (23)$$

Next, one can convert it to the  $d$ -to- $r_{pb}$ -converted effect size based on the equal-groups conversion. We also implemented a correction (see Equation 22) for the case in which only the point-biserial correlation between the experimental variable and continuous variable is already provided in a primary study, and the experimental and control group sizes are unequal. Note that we here assume that primary studies calculate the point-biserial correlation directly from the raw data and do not calculate Cohen's  $d$  and convert this  $d$  to a point-biserial correlation. Lastly, based on the study of de Jonge et al. (2020), we implemented an additional Cohen's  $d$  to biserial correlation ( $d$ -to- $r_b$ ) conversion, which can be used in other types of meta-analyses when there is a relation between an artificially dichotomized variable and continuous variable.

## Limitations and Suggestions for Future Research

One limitation of our study is that we only investigated the effect on the MASEM results in situations where the primary studies had quite similar characteristics. For example, when considering the effects of sample size, we only evaluated conditions in which all studies contained either small, medium, or large samples. This helped us to investigate the effect of within-study sample size on the results. In reality, meta-analyses may be applied to combinations of studies with large and small sample sizes. Future research may reveal the consequences of including a mix of sample sizes. However, we do not expect different results using the equal-groups conversion in such situations, because this conversion generally led to the most accurate estimates, which were generally unbiased, even when the sample sizes were small.

Another limitation of our study is that we cannot directly answer the question regarding the minimum number of primary studies, required sample sizes within each study, or the acceptable percentage of missing correlations necessary for conducting a reliable MASEM analysis. Based on the findings of this study, and consistent with previous research on MASEM (e.g., Jak & Cheung, 2020; Oort & Jak, 2016), we expect that reduced information—

such as fewer studies, smaller sample sizes, or more missing correlation—leads to less accurate MASEM parameter estimates. The objective of our study was to demonstrate the impact of several conversions on MASEM parameters, rather than to establish the specific strength of the bias in different scenarios. Therefore, answering the aforementioned question would require future simulation research.

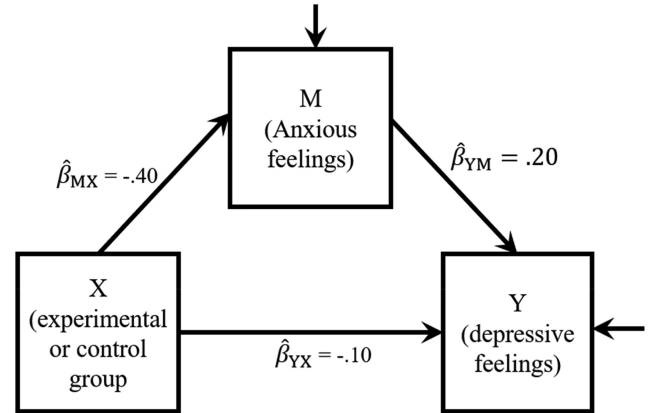
Although our study was not designed to systematically examine more extreme conditions, we did explore one such scenario as an extension of our main simulation study. Specifically, we examined the case in which all primary studies had an allocation ratio of 1:2. To this end, we ran 18 additional conditions, identical to those in the original study except that all primary studies had an allocation ratio of 1:2. The RBs for all the pooled correlations, path coefficients, and indirect effect are presented in S7 in the online supplemental materials. We generally observed similar patterns to those in the original simulation study for the pooled correlations, with the equal-groups conversion of Cohen's  $d$  to the point-biserial correlation continuing to perform best overall. However, for  $\hat{\beta}_{YX}$  and  $\hat{\beta}_{YM}$ , the equal-groups conversion of Cohen's  $d$  yielded RB values slightly exceeding 5% when the sample size was small (i.e., 26). We attribute this to the fact that path coefficients are complex, nonlinear functions of the pooled correlations. Although we still recommend converting Cohen's  $d$  to a point-biserial correlation using the equal-groups conversion for RCT data, as cautioned above, path coefficients and indirect effects are not guaranteed to be unbiased under other conditions than those investigated in our main simulation study.

Another possible limitation is that in this study, we focused on one research design, the randomized two-group postintervention design. In practice, one may want to include primary studies that follow other research designs. Such situations will certainly complicate the analysis because a different design may involve different ways of sampling, different expressions for Cohen's  $d$ , with a different meaning, different assumptions, different inferences, and different threats to internal validity. So, even though all studies may aim to answer the same research question, their effect sizes cannot always be directly aggregated in a meta-analysis. The different effect sizes should be theoretically logically combinable. This also implies, for example, that we do not advocate the inclusion of correlations between  $M$  and  $Y$  that are obtained from non-RCT studies. In addition, future research is needed to investigate the effect of including other research designs on MASEM results.

## Interpretational Challenges

In this study, we investigated the suitability of MASEM in analyzing RCT data using the R package metaSEM, while focusing on a simple partial mediation model. Despite the simplicity of this model, interpreting the results can be challenging. By focusing exclusively on RCT data, we avoided issues common in non-RCT studies, such as confounding effects (e.g., Valente et al., 2017). In RCTs, random assignment serves to eliminate initial differences between groups, which enhances internal validity and simplifies some interpretational aspects. However, the interpretation of the results remains complex for several reasons. For example, the model here pertains to the aggregated level, meaning effects are averages over subpopulations. In addition, some path coefficients represent (averages of) group differences established by the intervention (in our model  $\hat{\beta}_{MX}$  and  $\hat{\beta}_{YX}$ ), while others pertain to the relation between two variables representing individual differences (in our model  $\hat{\beta}_{YM}$ ). We here

**Figure 13**  
Example Mediation Model



provide guidance on interpreting MASEM results. Our starting point is that in intervention research, effects are often expressed as standardized mean differences (Cohen's  $d$  or Hedges'  $g$ ), while in MASEM, these are in terms of (standardized) path coefficients. Standardized path coefficients can be interpreted as standardized regression coefficients, ranging from  $-1$  (perfect negative relation) to  $0$  (no relation) to  $1$  (perfect positive relation). This also applies to indirect effects. Next, consider the hypothetical model in Figure 13, where variable  $X$  represents the intervention or group membership (being in the experimental condition, yes or no),  $M$  represents a continuous mediator (anxious feelings), and  $Y$  represents a continuous outcome variable (depressive feelings). In this model:

1.  $\hat{\beta}_{MX}$  indicates the relationship between group membership and anxious feelings. Here, the effect of the intervention is .40. This implies that—post intervention—in the total sample, 16% ( $.40^2 \times 100\%$ ) of the variance in anxious feelings can be ascribed to the intervention.
2.  $\hat{\beta}_{YX}$  represents the direct effect of the intervention on depressive feelings, independent of anxious feelings. The intervention has a  $-.10$  effect on depressive feelings over and above the effect of intervention on anxious feelings, explaining  $(-.10^2) \times 100\% = 1\%$  of the variance in depressive feelings.
3.  $\hat{\beta}_{YM}$  indicates the effect of anxious feelings on depressive feelings after controlling for group membership. Regardless of which group is considered, the effect of anxious feelings on depressive feelings equals  $\hat{\beta}_{YM}$  (i.e., .20), in line with the standard assumption in RCTs that the intervention results in mean differences in  $M$  and  $Y$  but does not affect the relation between  $M$  and  $Y$  itself (see the introductory part of the article).
4. Indirect effect ( $\hat{\beta}_{MX} \times \hat{\beta}_{YM}$ ) represents the effect of the intervention on depressive feelings through anxious feelings. In our example, this effect is .08 ( $-.40 \times .20$ ), meaning that 0.64% ( $.08^2 \times 100\%$ ) in depressive feelings is due to the intervention's effect on anxious feelings and the relation between anxious feelings and depressive feelings, hence on top of any direct effect that the intervention may have on depressive feelings.

All effects must be interpreted under the assumption that the fitted model is correct, implying the temporal order of  $X$ ,  $M$ , and  $Y$  is correct

and mediation is at least partial (meaning both  $\beta_{MX} \neq 0$  and  $\beta_{YM} \neq 0$ ) and possibly full (if, in addition,  $\beta_{YX} = 0$ ). Whether the effects reflect causal effects is a separate issue and beyond the scope of this study. We do stress that fitting a mediation model does not provide a test whether mediation is present but provides estimates of mediation effects under the hypothesis that the model is correct. Only falsification of causal effects is possible, verification is not.

## Conclusion

MASEM can be used to synthesize the results of multiple RCTs and investigate the indirect effect of an intervention variable on outcome variables. The typically reported statistics in RCTs must be converted before they can be included in the MASEM analysis. In general, we recommend converting Cohen's  $d$  to a point-biserial correlation using the equal-groups conversion (Equation 12); this generally resulted in the most accurate estimates of pooled correlations, path coefficients, and indirect effects (while standard errors were largely independent of conversion). Although our general recommendation is to convert Cohen's  $d$  to a point-biserial correlation using the equal-groups conversion, we caution that this does not guarantee unbiased path coefficients or indirect effects under other conditions than those examined in our simulation study.

The equal-groups conversion formula is not implemented in the mainstream conversion tools. To aid the meta-analysts, we developed a user-friendly web application entitled ESCACO (<https://hdejonge.shinyapps.io/ESCACO>) that converts the user's primary study statistics into an effect size suitable for use in MASEM.

## References

- Aaron, B., Kromrey, J. D., & Ferron, J. (1998). *Equating "r"-based and "d"-based effect size indices: problems with a commonly recommended formula*. Paper presented at the Annual Meeting of Florida Educational Research Association, Orlando, FL, United States. <https://eric.ed.gov/?id=ED433353>
- Agresti, A., & Finlay, B. (2014). *Statistical methods for the social sciences*. Pearson Education International.
- Becker, B. J. (1992). Using results from replicated studies to estimate linear models. *Journal of Educational Statistics*, 17(4), 341–362. <https://doi.org/10.2307/1165128>
- Becker, B. J. (1995). Corrections to "Using results from replicated studies to estimate linear models." *Journal of Educational and Behavioral Statistics*, 20(1), 100–102. <https://doi.org/10.2307/1165390>
- Ben-Shachar, M. S., Lüdtke, D., & Makowski, D. (2020). Effectsizes: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, 5(56), Article 2815. <https://doi.org/10.21105/joss.02815>
- Berger, T., Hohl, E., & Caspar, F. (2009). Internet-based treatment for social phobia: A randomized controlled trial. *Journal of Clinical Psychology*, 65(10), 1021–1035. <https://doi.org/10.1002/jclp.20603>
- Bijlsma, A. M., Assink, M., Stams, G. J. J., & van der Put, C. E. (2022). A critical evaluation of the risk, need, and responsivity principles in family interventions for delinquent youth: A meta-analysis. *Criminal Justice Review*, 49(3), 310–344. <https://doi.org/10.1177/07340168221140830>
- Bonett, D. G. (2020). Point-biserial correlation: Interval estimation, hypothesis testing, meta-analysis, and sample size determination. *British Journal of Mathematical and Statistical Psychology*, 73(Suppl. 1), 113–144. <https://doi.org/10.1111/bmsp.12189>
- Borenstein, M., & Hedges, L. V. (2019). Effect sizes for meta-analysis. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 207–245). Russell Sage Foundation.
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.
- Bretthauer, M., Løberg, M., Wieszczy, P., Kalager, M., Emilsson, L., Garborg, K., Rupinski, M., Dekker, E., Spaander, M., Bugajski, M., Holme, Ø., Zauber, A. G., Pilonis, N. D., Mroz, A., Kuipers, E. J., Shi, J., Hernán, M. A., Adami, H.-O., Regula, J., ... Kaminski, M. F. (2022). Effect of colonoscopy screening on risks of colorectal cancer and related death. *New England Journal of Medicine*, 387(17), 1547–1556. <https://doi.org/10.1056/NEJMoa2208375>
- Cai, Z., & Fan, X. (2020). A comparison of fixed-effects and random-effects models for multivariate meta-analysis using an SEM approach. *Multivariate Behavioral Research*, 55(6), 839–854. <https://doi.org/10.1080/00273171.2019.1689348>
- Cheung, M. W.-L. (2014). Fixed- and random-effects meta-analytic structural equation modeling: Examples and analyses in R. *Behavior Research Methods*, 46(1), 29–40. <https://doi.org/10.3758/s13428-013-0361-y>
- Cheung, M. W.-L. (2015a). *Meta-analysis: A structural equation modeling approach*. Wiley.
- Cheung, M. W.-L. (2015b). metaSEM: An R package for meta-analysis using structural equation modeling. *Frontiers in Psychology*, 5, Article 1521. <https://doi.org/10.3389/fpsyg.2014.01521>
- Cheung, M. W.-L. (2019). Some reflections on combining meta-analysis and structural equation modeling. *Research Synthesis Methods*, 10(1), 15–22. <https://doi.org/10.1002/jrsm.1321>
- Cheung, M. W.-L., & Chan, W. (2005). Meta-analytic structural equation modeling: A two-stage approach. *Psychological Methods*, 10(1), 40–64. <https://doi.org/10.1037/1082-989X.10.1.40>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Earlbaum Associates.
- de Jonge, H., Jak, S., & Kan, K. J. (2020). Dealing with artificially dichotomized variables in meta-analytic structural equation modeling. *Zeitschrift für Psychologie/Journal of Psychology*, 228(1), 25–35. <https://doi.org/10.1027/2151-2604/a000395>
- de Jonge, H., Kan, K. J., Oort, F., & Jak, S. (2024, July). *Project 'How to Synthesize Randomized Controlled Trial Data with Meta-Analytic Structural Equation Modeling: A Comparison of Various d-to-rpb Conversions'*. <https://osf.io/ynkj3>
- de Mooij, B., Fekkes, M., Scholte, R. H., & Overbeek, G. (2020). Effective components of social skills training programs for children and adolescents in nonclinical samples: A multilevel meta-analysis. *Clinical Child and Family Psychology Review*, 23(2), 250–264. <https://doi.org/10.1007/s10567-019-00308-x>
- Fernández-Castilla, B., Jamshidi, L., Declercq, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2020). The application of meta-analytic (multi-level) models with multiple random effects: A systematic review. *Behavior Research Methods*, 52(5), 2031–2052. <https://doi.org/10.3758/s13428-020-01373-9>
- Freeman, D., Sheaves, B., Goodwin, G. M., Yu, L. M., Nickless, A., Harrison, P. J., Emsley, R., Luik, A. I., Foster, R. G., Wadekar, V., Hinds, C., Gumley, A., Jones, R., Lightman, S., Jones, S., Bentall, R., Kinderman, P., Rowse, G., Brugha, T., ... Espie, C. A. (2017). The effects of improving sleep on mental health (OASIS): A randomised controlled trial with mediation analysis. *The Lancet Psychiatry*, 4(10), 749–758. [https://doi.org/10.1016/S2215-0366\(17\)30328-0](https://doi.org/10.1016/S2215-0366(17)30328-0)
- Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, 2(2), 156–168. <https://doi.org/10.1177/2515245919847202>
- Furrow, C. F., & Beretvas, S. N. (2005). Meta-analytic methods of pooling correlation matrices for structural equation modeling under different patterns of missing data. *Psychological Methods*, 10(2), 227–254. <https://doi.org/10.1037/1082-989X.10.2.227>
- Gabelica, M., Bojčić, R., & Puljak, L. (2022). Many researchers were not compliant with their published data sharing statement: A mixed-methods

- study. *Journal of Clinical Epidemiology*, 150, 33–41. <https://doi.org/10.1016/j.jclinepi.2022.05.019>
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *The Educational Researcher*, 5(10), 3–8. <https://doi.org/10.3102/0013189X005010003>
- Gnambs, T. (2023). A brief note on the standard error of the Pearson correlation. *Collabra: Psychology*, 9(1), Article 87615. <https://doi.org/10.1525/collabra.87615>
- Gnambs, T., & Sengewald, M. A. (2023). Meta-analytic structural equation modeling with fallible measurements. *Zeitschrift für Psychologie*, 231(1), 39–52. <https://doi.org/10.1027/2151-2604/a000511>
- Hamman, E. A., Pappalardo, P., Bence, J. R., Peacor, S. D., & Osenberg, C. W. (2018). Bias in meta-analyses using Hedges' *d*. *Ecosphere*, 9(9), Article e02419. <https://doi.org/10.1002/ecs2.2419>
- Hedges, L. V. (1981). Distribution theory for glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107–128. <https://doi.org/10.2307/1164588>
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.
- Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, 3(4), 486–504. <https://doi.org/10.1037/1082-989X.3.4.486>
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396), 945–960. <https://doi.org/10.1080/01621459.1986.10478354>
- Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling: An overview and a meta-analysis. *Sociological Methods & Research*, 26(3), 329–367. <https://doi.org/10.1177/0049124198026003003>
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage Publications.
- Jacobs, P., & Viechtbauer, W. (2017). Estimation of the biserial correlation and its sampling variance for use in meta-analysis. *Research Synthesis Methods*, 8(2), 161–180. <https://doi.org/10.1002/jrsm.1218>
- Jak, S. (2015). *Meta-analytic structural equation modelling*. Springer.
- Jak, S., & Cheung, M. W. L. (2020). Meta-analytic structural equation modeling with moderating effects on SEM parameters. *Psychological Methods*, 25(4), 430–455. <https://doi.org/10.1037/met0000245>
- Jak, S., Li, H., Kolbe, L., de Jonge, H., & Cheung, M. W.-L. (2021). Meta-analytic structural equation modeling made easy: A tutorial and web application for one-stage MASEM. *Research Synthesis Methods*, 12(5), 590–606. <https://doi.org/10.1002/jrsm.1498>
- Jansen, R. S., Van Leeuwen, A., Janssen, J., Jak, S., & Kester, L. (2019). Self-regulated learning partially mediates the effect of self-regulated learning interventions on achievement in higher education: A meta-analysis. *Educational Research Review*, 28, Article 100292. <https://doi.org/10.1016/j.edurev.2019.100292>
- Kline, T. J. (2005). *Psychological testing: A practical approach to design and evaluation*. Sage Publications.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, Article 863. <https://doi.org/10.3389/fpsyg.2013.00863>
- Leijten, P., Gardner, F., Melendez-Torres, G. J., Van Aar, J., Hutchings, J., Schulz, S., Knerr, W., & Overbeek, G. (2019). Meta-analyses: Key parenting program components for disruptive child behavior. *Journal of the American Academy of Child & Adolescent Psychiatry*, 58(2), 180–190. <https://doi.org/10.1016/j.jaac.2018.07.900>
- Lev, J. (1949). The point biserial coefficient of correlation. *The Annals of Mathematical Statistics*, 20(1), 125–126. <https://doi.org/10.1214/aoms/1177730103>
- Lin, L. (2018). Bias caused by sampling error in meta-analysis with small sample sizes. *PLOS ONE*, 13(9), Article e0204056. <https://doi.org/10.1371/journal.pone.0204056>
- Lin, L., & Aloe, A. M. (2020). Evaluation of various estimators for standardized mean difference in meta-analysis. *Statistics in Medicine*, 40(2), 403–426. <https://doi.org/10.1002/sim.8781>
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Russell Sage Foundation. <https://doi.org/10.1037/1082-989X.7.1.105>
- Lüdtke, D. (2019). *esc: Effect Size Computation for Meta Analysis* (Version 0.5.1). <https://doi.org/10.5281/zenodo.1249218>
- McGough, J. J., & Faraone, S. V. (2009). Estimating the size of treatment effects: Moving beyond *p* values. *Psychiatry (Edmont)*, 6(10), 21–29. <https://pmc.ncbi.nlm.nih.gov/articles/PMC2791668/>
- McGrath, R. E., & Meyer, G. J. (2006). When effect sizes disagree: The case of *r* and *d*. *Psychological Methods*, 11(4), 386–401. <https://doi.org/10.1037/1082-989X.11.4.386>
- Moeyaert, M., Ugille, M., Natasha Beretvas, S., Ferron, J., Bunuan, R., & Van den Noortgate, W. (2017). Methods for dealing with multiple outcomes in meta-analysis: A comparison between averaging effect sizes, robust variance estimation and multilevel meta-analysis. *International Journal of Social Research Methodology*, 20(6), 559–572. <https://doi.org/10.1080/13645579.2016.1252189>
- Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, 38(11), 2074–2102. <https://doi.org/10.1002/sim.8086>
- Ng, M. Y., DiVasto, K. A., Gonzalez, N.-A.-R., Cootner, S., Lipsey, M. W., & Weisz, J. R. (2023). How do cognitive behavioral therapy and interpersonal psychotherapy improve youth depression? Applying meta-analytic structural equation modeling to three decades of randomized trials. *Psychological Bulletin*, 149(9-10), 507–548. <https://doi.org/10.1037/bul0000395>
- Nissen, E. R., O'Connor, M., Kald, V., Højris, I., Borre, M., Zachariae, R., & Mehlsen, M. (2020). Internet-delivered mindfulness-based cognitive therapy for anxiety and depression in cancer survivors: A randomized controlled trial. *Psycho-Oncology*, 29(1), 68–75. <https://doi.org/10.1002/pon.5237>
- Olkin, I., & Siotani, M. (1976). Asymptotic distribution of functions of a correlation matrix. In S. Ikeda, et al. (Ed.), *Essays in probability and statistics*. Tokyo: Shinko Tsusho Co., Ltd.
- Oort, F. J., & Jak, S. (2016). Maximum likelihood estimation in meta-analytic structural equation modeling. *Research Synthesis Methods*, 7(2), 156–167. <https://doi.org/10.1002/jrsm.1203>
- Pustejovsky, J. E. (2014). Converting from *d* to *r* to *z* when the design uses extreme groups, dichotomization, or experimental control. *Psychological Methods*, 19(1), 92–112. <https://doi.org/10.1037/a0033788>
- R Core Team. (2022). *R: A language and environment for statistical computing*. <https://www.R-project.org>
- Salmela-Aro, K., Read, S., Rouhe, H., Halmesmäki, E., Toivanen, R. M., Tokola, M. I., & Saisto, T. (2012). Promoting positive motherhood among nulliparous pregnant women with an intense fear of childbirth: RCT intervention. *Journal of Health Psychology*, 17(4), 520–534. <https://doi.org/10.1177/1359105311421050>
- Sheng, Z., Kong, W., Cortina, J. M., & Hou, S. (2016). Analyzing matrices of meta-analytic correlations: Current practices and recommendations. *Research Synthesis Methods*, 7(2), 187–208. <https://doi.org/10.1002/jrsm.1206>
- Small, G. W., Siddarth, P., Li, Z., Miller, K. J., Ercoli, L., Emerson, N. D., Martinez, J., Wong, K.-P., Liu, J., Merrill, D. A., Chen, S. T., Henning, S. M., Satyamurthy, N., Huang, S.-C., Heber, D., & Barrio, J. R. (2018). Memory and brain amyloid and tau effects of a bioavailable form of curcumin in non-demented adults: A double-blind, placebo-controlled 18-month trial. *The American Journal of Geriatric Psychiatry*, 26(3), 266–277. <https://doi.org/10.1016/j.jagp.2017.10.010>
- Sobel, M. E. (1986). Some new results on indirect effects and their standard errors in covariance structure models. In N. B. Tuma (Ed.), *Sociological methodology* (pp. 159–186). Jossey-Bass.
- Tate, R. F. (1954). Correlation between a discrete and a continuous variable. Point-biserial correlation. *The Annals of Mathematical Statistics*, 25(3), 603–607. <https://doi.org/10.1214/aoms/1177728730>

- Tikuišis, R., Miliauskas, P., Samalavičius, N. E., Žurauskas, A., Samalavičius, R., & Zabulis, V. (2014). Intravenous lidocaine for post-operative pain relief after hand-assisted laparoscopic colon surgery: A randomized, placebo-controlled clinical trial. *Techniques in Coloproctology*, 18(4), 373–380. <https://doi.org/10.1007/s10151-013-1065-0>
- Valente, M. J., Pelham, W. E., III, Smyth, H., & MacKinnon, D. P. (2017). Confounding in statistical mediation analysis: What it is and how to address it. *Journal of Counseling Psychology*, 64(6), 659–671. <https://doi.org/10.1037/cou0000242>
- Van Erp, S., Verhagen, J., Grasman, R. P., & Wagenmakers, E. J. (2017). Estimates of between-study heterogeneity for 705 meta-analyses reported in *Psychological Bulletin* from 1990–2013. *Journal of Open Psychology Data*, 5(1), Article 4. <https://doi.org/10.5334/jopd.33>
- Van Zundert, C. H., & Miočević, M. (2020). A comparison of meta-methods for synthesizing indirect effects. *Research Synthesis Methods*, 11(6), 849–865. <https://doi.org/10.1002/jrsm.1445>
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). Springer.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>
- Viswesvaran, C., & Ones, D. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. *Personnel Psychology*, 48(4), 865–885. <https://doi.org/10.1111/j.1744-6570.1995.tb01784.x>
- Wang, X. S., Shi, Q., Mendoza, T., Lin, S., Chang, J. Y., Bokhari, R. H., Lin, H. K., Garcia-Gonzalez, A., Kamal, M., Cleeland, C. S., & Liao, Z. (2020). Minocycline reduces chemoradiation-related symptom burden in patients with non-small cell lung cancer: A phase 2 randomized trial. *International Journal of Radiation Oncology, Biology, Physics*, 106(1), 100–107. <https://doi.org/10.1016/j.ijrobp.2019.10.010>
- Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, 61(7), 726–728. <https://doi.org/10.1037/0003-066X.61.7.726>

## Appendix A

### Derivation of the $d$ -to- $r_{pb}$ Conversions

In general, a test statistic can be formulated as follows (e.g., Agresti & Finlay, 2014):

$$\text{Test statistic} = \frac{\text{parameter estimate} - \text{parameter value under } H_0}{\text{standard error of the parameter estimate}}. \quad (\text{A1})$$

The test statistic of a mean difference,  $\bar{y}_1 - \bar{y}_0$ , follows a  $t$  distribution with  $df = n_1 + n_0 - 2$ :

$$t = \frac{(\bar{y}_1 - \bar{y}_0) - (\mu_1 - \mu_0)}{SE_{\bar{y}_1 - \bar{y}_0}}. \quad (\text{A2})$$

Under the hypothesis that  $\mu_0 = \mu_1$ , and under the assumption that there is no effect in the population (i.e., the null hypothesis of equal means is true), this simplifies to:

$$t = \frac{(\bar{y}_1 - \bar{y}_0)}{SE_{\bar{y}_1 - \bar{y}_0}}. \quad (\text{A3})$$

Under the same assumption that there is no effect in the population, the test statistic of the point-biserial correlation ( $\rho_{pb} = 0$ ) also follows a  $t$  distribution with  $df = n_1 + n_0 - 2$  (Tate, 1954):

$$t = \frac{r_{pb} - 0}{SE_{r_{pb}}} = \frac{r_{pb}}{SE_{r_{pb}}}. \quad (\text{A4})$$

Provided with one and the same data set, the following equation then holds:

$$t = \frac{(\bar{y}_1 - \bar{y}_0)}{SE_{\bar{y}_1 - \bar{y}_0}} = \frac{r_{pb}}{SE_{r_{pb}}}. \quad (\text{A5})$$

The next step is to substitute the expressions of the standard error under the null hypothesis. For a correlation, this is (Cohen, 1988;

Gnambs, 2023):

$$SE_{r_{pb}} = \sqrt{\frac{1 - r_{pb}^2}{n_1 + n_0 - 2}}. \quad (\text{A6})$$

If the population variances in the two groups are equal (which is an assumption in true experimental designs), the expression of the standard error of the mean difference is:

$$SE_{\bar{y}_1 - \bar{y}_0} = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_0}}, \quad (\text{A7})$$

where  $s_p$  stands for the within-group pooled standard deviation:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_0 - 1)s_0^2}{n_1 + n_0 - 2}}, \quad (\text{A8})$$

and where  $s_0^2$  and  $s_1^2$  represent the within-group standard deviations in the two groups.

If the assumption of equal variances holds, Equation A5 becomes:

$$t = \frac{(\bar{y}_1 - \bar{y}_0)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_0}}} = \frac{r_{pb}}{\sqrt{\frac{1 - r_{pb}^2}{n_1 + n_0 - 2}}}. \quad (\text{A9})$$

As already shown in Equation 2, Cohen's  $d$  is defined as:

$$d = \frac{(\bar{y}_1 - \bar{y}_0)}{s_p}. \quad (\text{A10})$$

The equality in Equation A9 can thus be read as:

$$t = \frac{d}{\sqrt{\frac{1}{n_1} + \frac{1}{n_0}}} = \frac{r_{pb}}{\sqrt{\frac{1 - r_{pb}^2}{n_1 + n_0 - 2}}}. \quad (\text{A11})$$

(Appendices continue)

From this equation, it follows that  $d$  can be formulated in terms of  $r_{pb}$ :

$$d = \frac{r_{pb} \sqrt{\frac{1}{n_1} + \frac{1}{n_0}}}{\sqrt{\frac{1 - r_{pb}^2}{n_1 + n_0 - 2}}}, \quad (A12)$$

and, conversely,  $r_{pb}$  in terms of  $d$ :

$$r_{pb} = \frac{d \sqrt{\frac{1 - r_{pb}^2}{n_1 + n_0 - 2}}}{\sqrt{\frac{1}{n_1} + \frac{1}{n_0}}}. \quad (A13)$$

Simplifying the latter gives the formula provided by, for example, Jacobs and Viechtbauer (2017) and Lakens (2013):

$$r_{pb} = \frac{d}{\sqrt{d^2 + \frac{(n_1 + n_0)^2 - 2(n_1 + n_0)}{n_1 n_0}}}. \quad (A14)$$

Notably, in large samples, the  $t$  distribution will approach the standard normal  $z$  distribution such that the standard error approximates:

$$SE_{r_{pb}} \approx \sqrt{\frac{1 - r_{pb}^2}{n_1 + n_0}}. \quad (A15)$$

And the relationship between  $r_{pb}$  and  $d$  can be expressed (e.g., see Borenstein & Hedges, 2019; Borenstein et al., 2009; Lipsey & Wilson, 2001) as:

$$r_{pb} \approx \frac{d}{\sqrt{d^2 + \frac{(n_1 + n_0)^2}{n_1 n_0}}}. \quad (A16)$$

When group sizes are exactly equal ( $n_0 = n_1 = n = 1/2 N$ ), the relationship in Equation A14 simplifies to an expression that we rarely find in the literature, but see Aaron et al. (1998):

$$r_{pb} = \frac{d}{\sqrt{d^2 + 4 - \frac{8}{N}}}. \quad (A17)$$

In large samples, this will lead to the following equation, which does figure frequently in the literature (e.g., Borenstein & Hedges, 2019; Borenstein et al., 2009; Lipsey & Wilson, 2001):

$$r_{pb} = \lim_{n \rightarrow \infty} \frac{d}{\sqrt{d^2 + 4 - \frac{8}{N}}} = \frac{d}{\sqrt{d^2 + 4}}. \quad (A18)$$

## Appendix B

### Values of the Population Coefficients

This supplement explains the values of the population coefficients that we used in our simulation study. In line with the rationale behind randomized control trials (see the main text), we (a) split the infinite (super)population into two groups of equal sizes (still being infinite), meaning we assumed parameter  $\pi$  (see the main text) to be equal to .50, and (b) assumed that the intervention established a group mean difference in both  $M$  and  $Y$ , while the variances remained equal over the two groups. The group means on variable  $M$  were assumed to be  $\mu_{M,1} = .40$  and  $\mu_{M,0} = -.40$  and the within-group variance  $\sigma_{M,1}^2 = \sigma_{M,0}^2 = \sigma_M^2 = 1$ , which implies a standardized group mean difference in  $M$  of  $\delta_M = \frac{\mu_{M,1} - \mu_{M,0}}{\sigma_M} = \frac{.40 - (-.40)}{1} = .80$ , denoting a large effect according to Cohen's (1988) criteria. The total mean of variable  $M$  becomes  $\mu_{M,T} = \pi_1 \times \mu_{M,1} + \pi_2 \times \mu_{M,2} = .50 \times .40 + .50 \times -.40 = 0$ . Similarly, we assumed  $\mu_{Y,1} = .25$ ,  $\mu_{Y,0} = -.25$  and variance  $\sigma_{Y,1}^2 = \sigma_{Y,0}^2 = \sigma_Y^2 = 1$  such that  $\mu_{Y,T} = 0$  and  $\mu_{Y,T} = 0$  and  $\delta_Y = \frac{.25 - (-.25)}{1} = .50$ , denoting a medium effect (Cohen, 1988). Lastly, we assumed a preintervention correlation between  $M$  and  $Y$  of  $\rho = .40$ . Given the intervention only induces a shift in mean, this implies (1) a postintervention within-groups correlation of .40 (in both groups) and (2) total variances (i.e., the variance when groups are combined) of:

$$\begin{aligned}\sigma_{M,T}^2 &= \pi(\sigma_{M,1}^2 + (\mu_{M,1} - \mu_{M,T})^2) + (1 - \pi)(\sigma_{M,0}^2 + (\mu_{M,0} - \mu_{M,T})^2) \\ &= .50(1 + (.40 - 0)^2) + (1 - .50)(1 + (-.40 - 0)^2) = 1.16,\end{aligned}\quad (B1)$$

and

$$\begin{aligned}\sigma_{Y,T}^2 &= \pi(\sigma_{Y,1}^2 + (\mu_{Y,1} - \mu_{Y,T})^2) + (1 - \pi)(\sigma_{Y,0}^2 + (\mu_{Y,0} - \mu_{Y,T})^2) \\ &= .50(1 + (.25 - 0)^2) + (1 - .50)(1 + (-.25 - 0)^2) = 1.0625.\end{aligned}\quad (B2)$$

As a result, the total postintervention correlation between  $M$  and  $Y$ ,  $\rho_{YM}$ , becomes:

Equation B3 (see below)

Given the parameter values  $\pi = .50$ ,  $\delta_M = .80$ , and  $\delta_Y = .50$ , the point-biserial correlations in the population become (see Formula 4 in the main text):

$$\begin{aligned}\rho_{pb,MX} &= \frac{\delta_M}{\sqrt{\delta_M^2 + \frac{1}{\pi(1-\pi)}}} = \frac{.80}{\sqrt{.80^2 + \frac{1}{.50(1-.50)}}} \\ &= .3714,\end{aligned}\quad (B4)$$

and:

$$\begin{aligned}\rho_{pb,YX} &= \frac{\delta_Y}{\sqrt{\delta_Y^2 + \frac{1}{\pi(1-\pi)}}} = \frac{.50}{\sqrt{.50^2 + \frac{1}{.50(1-.50)}}} \\ &= .2425.\end{aligned}\quad (B5)$$

Given the path model (see Figure 2 in the main document) and the correlations  $\rho_{pb,MX} = .3714$ ,  $\rho_{pb,YX} = .2425$ , and  $\rho_{YM} = .4504$ , the path coefficients equal:

$$\begin{aligned}\beta_{MX} &= \rho_{pb,MX} = .3714, \\ \beta_{YX} &= \frac{\rho_{pb,MX}\rho_{YM} - \rho_{pb,YX}}{\rho_{pb,MX}^2 - 1} \\ &= \frac{.3713907 \times .4503773 - .2425356}{.3713907^2 - 1} = .0873, \text{ and} \\ \beta_{YM} &= \frac{\rho_{pb,MX}\rho_{pb,YX} - \rho_{YM}}{\rho_{pb,MX}^2 - 1} \\ &= \frac{.3713907 \times .2425356 - .4503773}{.3713907^2 - 1} = .4180.\end{aligned}\quad (B6)$$

The indirect effect of  $X$  on  $Y$  becomes:

$$\beta_{MX} \times \beta_{YM} = .3714 \times .4180 = .1552.\quad (B7)$$

Received February 21, 2024

Revision received June 13, 2025

Accepted July 7, 2025 ■

$$\begin{aligned}\rho_{YM} &= \frac{\pi(\rho + (\mu_{M,1} - \mu_{M,T})(\mu_{Y,1} - \mu_{Y,T})) + (1 - \pi)(\rho + (\mu_{M,0} - \mu_{M,T})(\mu_{Y,1} - \mu_{Y,T}))}{\sigma_{Y,T}\sigma_{M,T}} \\ &= \frac{.50(.40 + (.40 - 0)(.25 - 0)) + (1 - .50)(.40 + (-.40 - 0)(-.25 - 0))}{\sqrt{1.16}\sqrt{1.0625}} = .4504.\end{aligned}\quad (B3)$$