

Meta-analyzing non-preregistered and preregistered studies

Robbie C.M. van Aert<sup>1</sup>

<sup>1</sup> Department of Methodology and Statistics, Tilburg University

Author Note

The authors made the following contributions. Robbie C.M. van Aert:  
Conceptualization, Formal analysis, Methodology, Software, Writing - Original Draft  
Preparation, Writing - Review & Editing.

Correspondence concerning this article should be addressed to Robbie C.M. van Aert,  
P.O. Box 90153, 5000 LE Tilburg, The Netherlands. E-mail:  
R.C.M.vanAert@tilburguniversity.edu

## Abstract

Preregistration is gaining ground in psychology, and a consequence of this is that preregistered studies are more often included in meta-analyses. Preregistered studies likely mitigate the effect of publication bias in a meta-analysis, because preregistered studies can be located in the registries they were registered in even if they do not get published. However, current meta-analysis methods do not take into account that preregistered studies are less susceptible to publication bias. Traditional methods treat all studies as equivalent while meta-analytic conclusions can be improved by taking advantage of preregistered studies. The goal of this paper is to introduce the Hybrid Extended Meta-Analysis (HYEMA) method that takes into account whether a study is preregistered or not to correct for publication bias in only the non-preregistered studies. The proposed method is applied to two meta-analyses on prominent effects in the psychological literature: the red-romance hypothesis and money priming. Applying HYEMA to these meta-analyses shows that the average effect size estimate is substantially closer to zero than the estimate of the random-effects meta-analysis model. Two simulation studies tailored to the two applications are also presented to illustrate the method's superior performance compared to the random-effects meta-analysis model and PET-PEESE when publication bias is present. Hence, I recommend to apply HYEMA as a sensitivity analysis if a mix of both preregistered and non-preregistered studies are present in a meta-analysis. R code as well as a web application (<https://rcmvanaert.shinyapps.io/HYEMA>) have been developed and are described in the paper to facilitate application of the method.

*Keywords:* meta-analysis, preregistration, publication bias, red-romance hypothesis, money priming

Word count: 6763

Meta-analyzing non-preregistered and preregistered studies

## Introduction

There is an uptake of preregistration in psychological science in particular and the social sciences in general. An exponential increase in the number of preregistered studies is observed that had been preregistered on the Open Science Framework (OSF) and AsPredicted with approximately 5,000, 10,000, and 40,000 preregistrations in 2016, 2018, and 2020, respectively (Nosek et al., 2022). Other evidence for the uptake of preregistration is that between 27% and 57% of surveyed psychologists indicated that they had conducted a preregistration (Makel, Hodges, Cook, & Plucker, 2021; van den Akker, Scherer, Wicherts, & Koole, 2020; Washburn et al., 2018), and 43% of studies published in top journals in psychology in 2022 contained at least one preregistered study (Simonsohn, 2023).

Preregistered studies have favorable properties that make these studies substantially different from non-preregistered conventional studies (henceforth conventional studies) and make them less likely at risk of bias. A first favorable property is that there is a clear distinction in preregistered studies, in contrast to conventional studies, between confirmatory and exploratory research (Nosek, Ebersole, DeHaven, & Mellor, 2018; Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012). This distinction informs readers about which analyses have been planned prior to and after the data collection. Fixing the research design and data analysis before the data collection restricts researchers to follow the prespecified procedures and prevent them from using researchers degrees of freedom to get desirable results (a.k.a. *p*-hacking, Simmons, Nelson, & Simonsohn, 2011; Wicherts et al., 2016). Another favorable property is that preregistered studies can be identified in registries that would otherwise not be readily accessible for researchers (Chambers & Munafò, 2013; Nosek & Lindsay, 2018). This will likely decrease the overestimation of effect size, because the effect of publication bias, the selective publication of results with favorable outcomes (e.g., Rothstein, Sutton, & Borenstein, 2005), is diminished. Even if publication bias hinders a

study from being published, it can still be located from the registry it was preregistered in, and information about the study may be obtained by contacting the authors. Ensinnck and Lakens (2023) showed that only 104 (61.5%) out of a random sample of 169 preregistrations on the OSF were published four years after the preregistration date. This implies that 65 (38.5%) of these studies could most likely only be located via its preregistration.

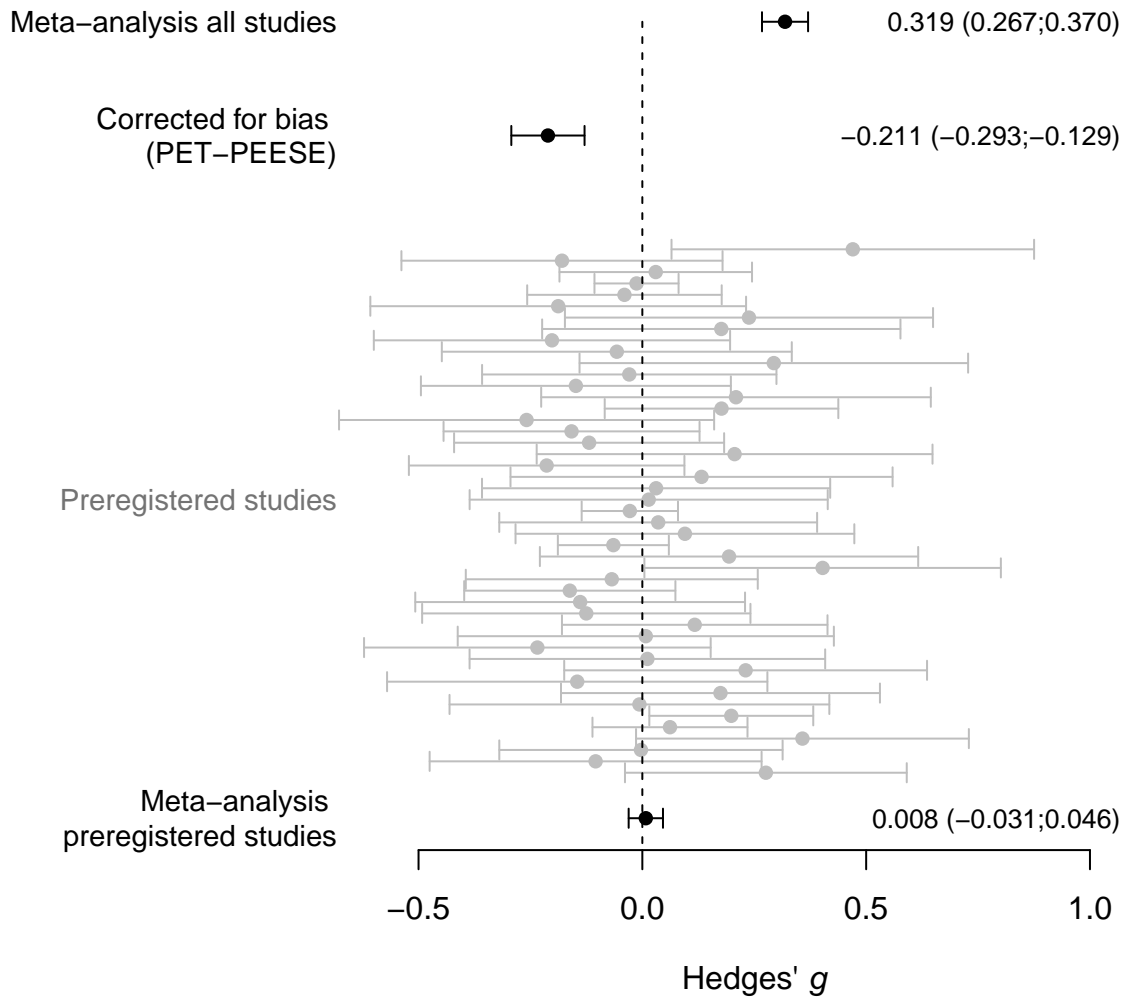
The increased number of preregistrations implies that more and more preregistered studies can be included in a meta-analysis to statistically combine effect sizes of multiple studies (Borenstein, Hedges, Higgins, & Rothstein, 2009, Chapter 1). Preregistered studies are treated the same as conventional studies in meta-analyses. That is, the traditional random-effects and equal-effect (a.k.a. common-effect or fixed-effect) meta-analysis models assume that the effect size estimates of all included (conventional and preregistered) studies are unbiased (Jackson & White, 2018; Viechtbauer, 2021). However, there is evidence that the conventional studies in the literature are biased (e.g., Bartoš et al., 2023; Camerer et al., 2016; Errington et al., 2021; Franco, Malhotra, & Simonovits, 2014; Open Science Collaboration, 2015). Applying the traditional meta-analysis models for synthesizing conventional and preregistered studies is suboptimal, because conventional studies are more likely at risk of bias compared to preregistered studies and can therefore yield substantially different effect sizes. Traditional meta-analysis models do not draw a distinction between conventional and preregistered studies and treat all studies equally. All studies are also treated equally in publication bias methods that meta-analysts commonly apply to test and correct for the presence of publication bias. That is, publication bias methods assume that all included studies in the meta-analysis were susceptible for bias. However, publication bias is most likely stronger in conventional compared to preregistered studies.

A possible analysis approach where not all studies are treated equally using traditional meta-analysis models and publication bias methods is to include a dichotomous moderator to study the difference in effect size between the conventional and preregistered studies. A

drawback of this approach is that it does not estimate one overall meta-analytic effect; it estimates whether the effect of, for example, the conventional studies is larger than of the preregistered studies. Furthermore, there is either not corrected for publication bias in all studies with traditional meta-analysis models or there is corrected for publication bias in all studies with publication bias methods.

An example where traditional meta-analysis models and publication bias methods are likely suboptimal for the analysis is the meta-analysis by Lodder, Ong, Grasman, and Wicherts (2019) that will be used as one of the examples throughout this paper. This meta-analysis is about the effect of money priming and includes 236 studies of which 47 (19.9%) were preregistered (see section “Money priming” for more details). Figure 1 provides a summary of the results of this meta-analysis. The meta-analytic estimate is 0.319 and statistically significant ( $z=12.133$ ,  $p<.001$ ). The estimate corrected for bias using the PET-PEESE method (Stanley & Doucouliagos, 2014) is -0.211 and also statistically significant ( $t(234)=-5.080$ ,  $p<.001$ ). PET-PEESE corrects for small-study effects and publication bias is one of the causes of small-study effects (for other causes see Egger, Smith, Schneider, & Minder, 1997). The preregistered studies are also shown in Figure 1, and the vast majority of the effect sizes (87.2%) of these preregistered studies are in between the estimates of the meta-analysis and PET-PEESE. This raises the question: What can be concluded for this meta-analysis given the divergent results? The meta-analytic estimate is most likely overestimated due to the likely bias in the conventional studies whereas PET-PEESE might overcorrect for bias since it assumes that the preregistered studies are also affected by bias.

The goal of this paper is to introduce the first meta-analysis method that treats conventional and preregistered studies differently and only corrects for publication bias in the conventional studies. The proposed Hybrid Extended Meta-Analysis (HYEMA) method yields more accurate effect size estimation and statistical inference than methods that do not



*Figure 1.* Summary of the results of the meta-analysis on money priming (Lodder et al., 2019). The numbers at the right-hand side of the figure are the average effect size estimates and 95% confidence intervals (in brackets).

take into account whether a study is preregistered or not. Hence, HYEMA is recommended to be applied as a sensitivity analysis in meta-analyses that contain a mix of conventional and preregistered studies.

HYEMA is related to several existing methods. The method is an extension of the hybrid meta-analysis method (van Aert & van Assen, 2018) for combining a single conventional and a single (preregistered) replication. HYEMA is an improvement in three ways: (1) it allows for the inclusion of multiple conventional and multiple preregistered studies, (2) heterogeneity in true effect sizes is estimated in the model, and (3) statistically nonsignificant conventional studies can be included. The proposed method is also related to selection model approaches to correct for publication bias in a meta-analysis such as the  $p$ -uniform (van Aert, Wicherts, & van Assen, 2016; van Assen, van Aert, & Wicherts, 2015) and  $p$ -uniform\* method (van Aert & van Assen, 2024) and the step weight function selection model (Hedges, 1984; Iyengar & Greenhouse, 1988; Vevea & Hedges, 1995) that is also called the three parameter selection model in the literature (Carter, Schönbrodt, Gervais, & Hilgard, 2019; McShane, Böckenholt, & Hansen, 2016). HYEMA is related to these methods, because correcting for bias in the conventional studies is done in a similar way. HYEMA is also related to a Bayesian approach (Carter & McCullough, 2018) where the focus is on whether prior beliefs based on the results of a meta-analysis shift when these prior beliefs are combined with high-quality preregistered replications.

The remainder of this paper is structured as follows. The next section describes the random-effects meta-analysis model that does not distinguish conventional from preregistered studies. HYEMA is then introduced and applied to meta-analyses on the red-romance hypothesis (Lehmann, Elliot, & Calin-Jageman, 2018) and money priming (Lodder et al., 2019). Next, simulation studies are performed tailored to the two applications to illustrate that HYEMA outperforms the random-effects meta-analysis model if publication bias is present in the conventional studies. The paper concludes with a discussion section.

### The random-effects meta-analysis model

The random-effects meta-analysis model can be written as (Konstantopoulos & Hedges, 2019)

$$y_i = \mu + \zeta_i + \epsilon_i$$

where  $y_i$  is the observed effect size of the  $i = 1, 2, \dots, k^{th}$  study,  $\mu$  is the average true effect size,  $\zeta_i$  is the study specific random effect that denotes the difference between  $\mu$  and a study's true effect size, and  $\epsilon_i$  is the sampling error of the  $i^{th}$  study. The random effects  $\zeta_i$  are assumed to follow a normal distribution with mean 0 and variance  $\tau^2$ . The parameter  $\tau^2$  indicates the between-study variance in true effect size, which is also known as the heterogeneity in the random-effects model. If  $\tau^2 = 0$ , the random-effects model simplifies to the equal-effect meta-analysis model. With respect to the sampling errors it is assumed that  $\epsilon_i \sim N(0, \sigma_i^2)$  where  $\sigma_i^2$  is the within-study sampling variance that is estimated and then assumed to be known. Mutual independence is assumed for all  $\zeta_i$  and  $\epsilon_i$ .

### The hybrid extended meta-analysis method

The random-effects model does not distinguish conventional from preregistered studies, and uses the same likelihood function for both types of studies. If publication bias is present, statistically significant studies are overrepresented in the meta-analysis and yield a positive bias in the estimator of the average effect. HYEMA treats studies differently depending on whether a study is a conventional or a preregistered study. It treats studies differently by using distinct likelihood functions for both types of studies. By using distinct likelihood functions, HYEMA takes into account that statistically significant studies are likely overrepresented in the meta-analysis due to publication bias. This yields less biased estimation and more accurate statistical inference of HYEMA compared to the random-effects model.

The likelihood function of a conventional study in HYEMA is the probability density



function of the truncated normal distribution. This likelihood function is a conditional likelihood depending on whether the effect size is statistically significant or nonsignificant using a right-tailed hypothesis test,

$$L_C(\mu, \tau^2; y_i, \sigma_i^2, y_i^{cv}) = \begin{cases} \frac{1}{\sqrt{\sigma_i^2 + \tau^2}} \phi\left(\frac{y_i - \mu}{\sqrt{\sigma_i^2 + \tau^2}}\right) & \text{if } p_i \leq \alpha \\ \frac{1 - \Phi\left(\frac{y_i^{cv} - \mu}{\sqrt{\sigma_i^2 + \tau^2}}\right)}{\frac{1}{\sqrt{\sigma_i^2 + \tau^2}} \phi\left(\frac{y_i - \mu}{\sqrt{\sigma_i^2 + \tau^2}}\right)} & \text{if } p_i > \alpha \end{cases} \quad (1)$$

where  $\phi$  and  $\Phi$  refer to the probability density and cumulative distribution function of the standard normal distribution,  $p_i$  is the right-tailed  $p$ -value of the  $i^{th}$  study,  $\alpha$  is the alpha-level for testing the null-hypothesis of no effect in the conventional studies, and  $y_i^{cv}$  is the  $(1 - \alpha)$ -quantile of a normal distribution with mean 0 and variance  $\sigma_i^2$  (i.e.,  $y_i^{cv}$  is the threshold that defines whether a conventional study is statistically significant or not).

Estimating the parameters with the conditional likelihood function in equation (1) corrects for publication bias in the conventional studies. That is, the conditional likelihood computes how likely observing an effect size is given that the effect size is statistically (non)significant, and statistical significance is also a criterion that affects the publication probability due to publication bias. Note that the likelihood function is presented for a situation where statistically significant positive effect sizes are more likely to be published than other effect sizes, because a right-tailed hypothesis test is assumed in the likelihood of equation (1). However, the conditional likelihood function can also be accommodated to take into account that significant negative rather than significant positive effect sizes are more likely to be published in case of publication bias.

The likelihood function of the preregistered studies is an *unconditional* likelihood function that is equal to the probability density function of the normal distribution.

Statistical significance of the preregistered study does not have to be taken into account, because it is assumed that these studies are accessible to the meta-analyst irrespective of their significance. The likelihood function of a preregistered study is

$$L_P(\mu, \tau^2; y_i, \sigma_i^2) = \frac{1}{\sqrt{\sigma_i^2 + \tau^2}} \phi\left(\frac{y_i - \mu}{\sqrt{\sigma_i^2 + \tau^2}}\right). \quad (2)$$

Let  $\mathbb{1}_i$  be an indicator function that is 1 in case the  $i^{th}$  study is a conventional study and 0 if it is a preregistered study. The likelihood of the  $i^{th}$  study is then equal to

$$\prod_{i=1}^k \left( L_C(\mu, \tau^2; y_i, \sigma_i^2, y_i^{cv}) \times \mathbb{1}_i + L_P(\mu, \tau^2; y_i, \sigma_i^2) \times (1 - \mathbb{1}_i) \right),$$

and parameter estimates for  $\mu$  and  $\tau^2$  can be obtained by maximizing the likelihood function.

## Statistical inference

Standard errors of the estimated parameters are obtained by taking the square root of the diagonal elements of the inverted Hessian (Pawitan, 2013, sec. 2.7). The obtained standard errors can be used for conducting Wald-type hypothesis tests and creating 95% confidence intervals. Another option for hypothesis testing is to use the likelihood-ratio test where the log-likelihood of a model with one of the two parameters constrained to zero is compared to the model where both parameters are unconstrained (e.g., Casella & Berger, 2002, sec. 8.2.1). A 95% confidence interval can also be computed based on the likelihood-ratio statistic by searching for the interval of the parameter where the likelihood-ratio test is not rejected given a particular  $\alpha$ -level (e.g., Agresti, 2013, sec. 3.1.8; Pawitan, 2013, sec. 2.6).

Throughout the paper, the Wald-type hypothesis test and confidence interval are reported for inferences for  $\mu$ , and the likelihood-ratio test and corresponding confidence interval are reported for inferences for  $\tau^2$ . The reason for making inference for  $\tau^2$  based on the likelihood-ratio test is that this is recommended for testing variance components

(Molenberghs & Verbeke, 2007)<sup>1</sup>. Furthermore, the likelihood-ratio test also better controlled the Type-I error rate and had larger statistical power than the Wald-test when testing for no between-study variance in a meta-analysis (Viechtbauer, 2007b).

## Applications

HYEMA, the random-effects model, and PET-PEESE are illustrated using data of two meta-analyses in the psychology literature. HYEMA was compared to the random-effects model and PET-PEESE, because the latter two do not correct or correct for bias in all studies included in the meta-analysis. The methods were applied to meta-analyses on the red-romance hypothesis (Lehmann et al., 2018) and money priming (Lodder et al., 2019). These two meta-analyses were selected, because they study prominent effects in psychological science and both examples contain conventional as well as preregistered studies. Furthermore, there is evidence that publication bias is present in the literature on the red-romance hypothesis and money priming (e.g., Francis, 2013; Vadillo, Hardwicke, & Shanks, 2016). Both meta-analyses used Hedges'  $g$  as effect size measure, but HYEMA can be applied to any effect size measure whose sampling distribution can be assumed to be normally distributed (e.g., raw mean difference, [Fisher's  $z$  transformed] correlations, log odds ratio).

R (R Core Team, 2023, Version 4.3.3) was used for the analyses. The R packages “puniform” (van Aert, 2023, Version 0.2.7) and “metafor” (Viechtbauer, 2010, Version 4.6.0) were used for applying HYEMA and fitting the random-effects model, respectively. The data of the meta-analysis on money priming and R code of these analyses is available at <https://osf.io/fhpzq/>. The data of the meta-analysis on the red-romance hypothesis were obtained from the R package “metadat” (White, Noble, Senior, Hamilton, & Viechtbauer,

---

<sup>1</sup>  $P$ -values of the likelihood-ratio test for testing  $H_0 : \tau^2 = 0$  are obtained by using  $0.5(\chi^2(0) + \chi^2(1))$  as reference distribution. The test statistic asymptotically follows this mixture distribution, because the tested hypothesis lies on the boundary of the parameter space (Andrews, 2001; Molenberghs & Verbeke, 2007; Viechtbauer, 2007b).

2022, Version 1.2.0).

## Red-romance hypothesis

The red-romance hypothesis implies that “the color red enhances heterosexual attraction in romantic contexts” (Lehmann et al., 2018, p. 1). All studies included in the meta-analysis are experiments where participants were shown pictures of a person of the opposite gender. The experimental manipulation was the exposure to the color red that was shown in the picture. In the control condition, participants were exposed with white, gray, green, or blue instead of red. The dependent variable of interest was the perceived attractiveness of the person in the picture. Lehmann et al. (2018) conducts two separate meta-analyses for male participants rating women and female participants rating men. In this section, the results are reported for the analyses of studies where female participants are rating men, because a larger portion of these studies was preregistered. The results of the analyses of studies where male participants are rating women are available via <https://osf.io/jfd42>.

This meta-analysis consists of 36 studies, and 8 (22.2%) of these studies were preregistered. The results of fitting the random-effects model<sup>2</sup> are presented in the first row of Table 1. The average effect size estimate was 0.128 and was significantly different from zero ( $z=2.431$ ,  $p=.015$ ). The estimated between-study variance was 0.044, and was also statistically significant ( $Q(35)=73.048$ ,  $p<.001$ ). The second row of Table 1 presents the results of PET-PEESE. The average effect size estimate corrected for bias with PET-PEESE

---

<sup>2</sup> In this section, the maximum likelihood estimator (e.g., Hardy & Thompson, 1996) is used for estimating the between-study variance in true effect size in the random-effects model. This is different from the restricted maximum likelihood estimator that was used in the meta-analyses on the red-romance hypothesis and money priming. The reason for using the maximum likelihood estimator was that HYEMA also uses the maximum likelihood estimator, so this made the results of the two methods directly comparable. A consequence of using different estimators was that the reported results of the random-effects model slightly deviated from the ones reported in Lehmann et al. (2018) and Lodder et al. (2019).

(-0.127) was of about the same size as the estimate of the random-effects model but the estimate was negative rather than positive. The null-hypothesis of no effect was not rejected with PET-PEESE ( $t(34)=-1.003$ ,  $p=.323$ ). PET-PEESE is based on weighted least squares regression and does not estimate and provide statistical inference for the between-study variance parameter.

The random-effects model was also fitted separately to the conventional and preregistered studies. These results show that the average effect size estimate differed more from zero based on the conventional studies (0.207) compared to the estimate of the preregistered studies (-0.097), and the average effect size was significantly different from zero for the conventional ( $z=3.183$ ,  $p=.001$ ) but not for the preregistered studies ( $z=-1.349$ ,  $p=.177$ ).

Table 1

*Results of applying the random-effects meta-analysis (RE MA) model, PET-PEESE, and HYEMA to the meta-analyses about the red-romance hypothesis (Lehmann et al., 2018) and money priming (Lodder et al., 2019).*

	$\hat{\mu}$ (SE)	(95% CI)	$H_0: \mu=0$	$\hat{\tau}^2$ (SE)	(95% CI)	$\hat{\tau}$	$H_0: \tau^2=0$
<b>Red-romance hypothesis</b>							
RE MA	0.128 (0.052)	(0.025;0.230)	$z=2.431, p=.015$	0.044 (0.022)	(0.021;0.175)	0.210	$Q(35)=73.048, p<.001$
PET-PEESE	-0.127 (0.127)	(-0.385;0.131)	$t(34)=-1.003, p=.323$	-	-	-	-
HYEMA	0.018 (0.047)	(-0.074;0.110)	$z=0.382, p=.703$	0.009 (0.014)	(0.000;0.052)	0.095	$LR=0.683, p=.204$
<b>Money priming</b>							
RE MA	0.319 (0.026)	(0.267;0.370)	$z=12.133, p<.001$	0.118 (0.015)	(0.102;0.171)	0.344	$Q(235)=1022.133, p<.001$
PET-PEESE	-0.211 (0.042)	(-0.293;-0.129)	$t(234)=-5.080, p<.001$	-	-	-	-
HYEMA	0.176 (0.032)	(0.112;0.239)	$z=5.435, p<.001$	0.080 (0.012)	(0.059;0.107)	0.283	$LR=233.900, p<.001$

*Note:*

PET-PEESE is based on weighted least squares regression and does not estimate or provide statistical inference for  $\tau^2$ . SE = Standard Error; CI = Confidence Interval; LR = Likelihood Ratio. The confidence interval of  $\hat{\tau}^2$  is computed with the  $Q$ -profile method (Viechtbauer, 2007a).

The first row of Figure 2 shows the contour-enhanced funnel plots (Peters, Sutton, Jones, Abrams, & Rushton, 2008) of the conventional (left panel) and preregistered studies (right panel). Effect sizes of studies (i.e., black circles) inside the white-shaded area are not statistically significant when testing a two-tailed hypothesis with  $\alpha=0.05$ . Other shaded areas indicate that the  $p$ -value of an effect size is smaller than 0.05. The farther away the effect size is from the white-shaded area, the smaller is the  $p$ -value. These funnel plots show that 10 of the 28 conventional studies (35.7%) were statistically significant. None of the preregistered studies were statistically significant.

The funnel plots can also be used to assess whether there are small-study effects, because an asymmetric funnel plot is evidence for the presence of these effects (Light & Pillemer, 1984). Effect sizes seem to be missing in the lower left part of the funnel plot of conventional studies making the funnel plot asymmetric. This was also confirmed by testing the null-hypothesis of no small-study effects. The dashed black line in the funnel plots reflects the estimated regression equation in Egger's regression test (Egger et al., 1997). This line is supposed to be a vertical line in case no small-study effects are present. The line in the plot of the conventional studies was not vertical, and this was confirmed by a statistically significant Egger's regression test ( $z=3.084$ ,  $p=.002$ ). The funnel plot of the preregistered studies was symmetric with an almost vertical line, and Egger's regression test was not statistically significant ( $z=-0.092$ ,  $p=.927$ ). This meta-analytic evidence illustrates that conventional and preregistered studies yield substantially different results, and that applying HYEMA as a sensitivity analysis is warranted.

The results of HYEMA are presented in the third row of Table 1. The average effect size estimate of HYEMA was substantially smaller than of the random-effects model (0.018 vs. 0.128) and close to zero. The average effect size was also no longer significantly different from zero ( $z=0.382$ ,  $p=.703$ ). The same pattern was observed for the between-study variance. The estimated between-study variance decreased from 0.044 in the random-effects model to

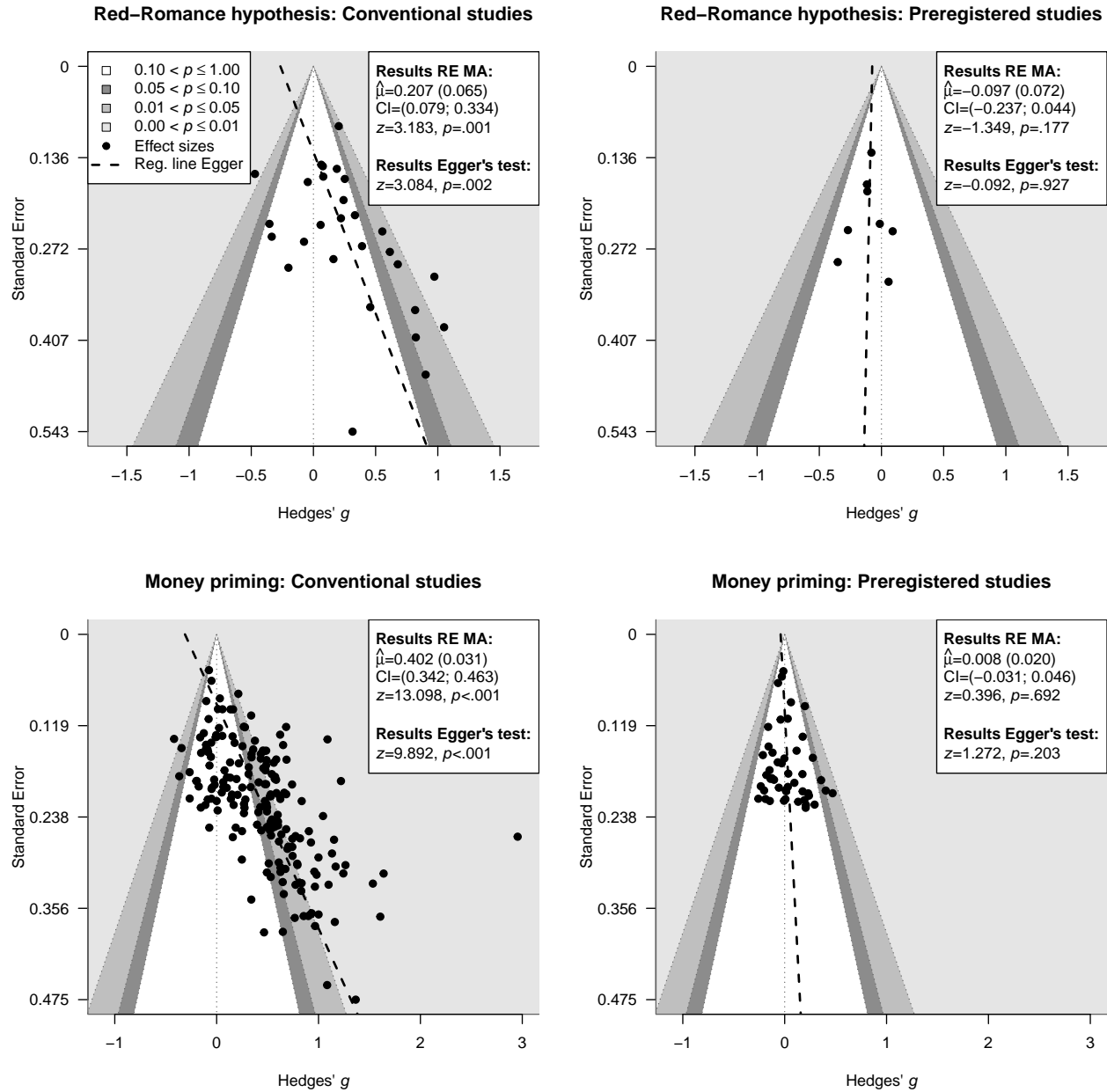


Figure 2. Contour-enhanced funnel plots split by whether a study was preregistered or not for the meta-analyses about the red-romance hypothesis (Lehmann et al., 2018, first row) and money priming (Lodder et al., 2019, second row). Reg. line Egger = the regression line of Egger's regression test.



0.009 when HYEMA was used, but the estimation was imprecise given the wide confidence interval  $([0.000; 0.052])$ . The null-hypothesis of no heterogeneity was also not rejected with HYEMA ( $LR=0.683$ ,  $p=.204$ ).

## Money priming

Lodder et al. (2019) conducted a meta-analysis on the effect of money priming, which refers to the effect that people take a more self-sufficient orientation or behave in a more self-sufficient way if they receive a money-related manipulation (Vohs, Mead, & Goode, 2006). The included studies in the meta-analysis typically contain an experimental and control condition where the participant is only primed with money in the experimental condition. After the manipulation, participants' self-sufficient orientation or behavior is measured, and a larger average score in the experimental than in the control group is interpreted as evidence in favor of the existence of a money priming effect.

The results of fitting the random-effects model and PET-PEESE to data of all studies were already described in the introduction of this paper and are presented in the last three rows of Table 1. Conducting separate meta-analyses for the conventional and preregistered studies showed that the average effect size estimate of the random-effects model was substantially larger for the conventional compared to the preregistered studies (0.402 vs. 0.008) and significantly different from zero based on the conventional studies ( $z=13.098$ ,  $p<.001$ ) but not when based on the preregistered studies ( $z=0.396$ ,  $p=.692$ ).

The second row of Figure 2 shows the contour-enhanced funnel plots separately for the conventional (left panel) and preregistered studies (right panel). A large portion of the conventional studies was statistically significant (109, 57.7%) compared to only 3 (6.4%) of the preregistered studies. There is evidence for the presence of small-study effects in the subset of conventional studies (Egger's test:  $z=9.892$ ,  $p<.001$ ) but not in the subset of preregistered studies (Egger's test:  $z=1.272$ ,  $p=.203$ ). These results indicate that type of

study should be taken into account in the analysis.

The sixth row of Table 1 shows the results of applying HYEMA to the meta-analysis on money priming. The average effect size estimate almost halved (0.176) compared to the estimate of the random-effects model (0.319), but the null-hypothesis of no effect was still rejected ( $z=5.435$ ,  $p<.001$ ). The estimate of the between-study variance was lower than estimated in the random-effects model (0.080 vs. 0.118), but it remained statistically significant ( $LR=233.900$ ,  $p<.001$ ).

### Simulation studies

The two applications above have shown that applying HYEMA can yield different conclusions than the random-effects model and PET-PEESE. Two simulation studies tailored towards the characteristics of the two applications are presented in this section to examine the statistical properties of HYEMA in comparison with the random-effects model and PET-PEESE.

### Method

The effect size estimate of the  $i^{th}$  study was generated using the marginal distribution of the random-effects model,  $y_i \sim N(\mu, \sigma_i^2 + \tau^2)$ . In case of a preregistered study, the effect size estimate was stored, and the next effect size estimate was generated. In case of a conventional study, there was first tested whether the null-hypothesis of no effect was rejected using a right-tailed test with  $\alpha=0.025$ . This resembles the common practice in psychology of testing two-tailed hypotheses where effect sizes in one direction are expected. The effect size of a conventional study was always included in the meta-analysis if it was statistically significant. Publication bias was introduced for a conventional study by including a statistically nonsignificant effect size in the meta-analysis with the probability  $pub$ . This was implemented by generating a value from a uniform distribution ranging from 0 to 1, and including a nonsignificant effect size if the value drawn from the uniform distribution was

larger than *pub*. If the generated effect size was not included in the meta-analysis, a new effect size of a conventional study was generated using the outlined procedure.

The design of the simulation studies was tailored towards the two applications by using the same number of conventional and preregistered studies as observed in the applications. Furthermore, the within-study sampling variances (i.e.,  $\sigma_i^2$ ) of the studies in the applications were used in the simulations such that the generated effect sizes of the conventional and preregistered studies were based on identical sampling variances as those in the applications. Conditions for  $\mu$  and  $\tau^2$  were partly based on the parameter estimates of the random-effects model and HYEMA. That is, data for the red-romance hypothesis were generated with  $\mu=0.128$  in combination with  $\tau^2=0.044$  and  $\mu=0.018$  in combination with  $\tau^2=0.009$ , and for money priming with  $\mu=0.319$  in combination with  $\tau^2=0.118$  and  $\mu=0.176$  in combination with  $\tau^2=0.080$ . The condition  $\mu = 0$  in combination with both estimates of  $\tau^2$  for the two meta-analyses were also included to enable studying the Type-I error rate. The conditions for *pub* were the same for both simulation studies and were equal to 0, 0.5, 0.9, and 1.

HYEMA was compared to the random-effects model and PET-PEESE in the simulations. The random-effects model was included as comparison where there is no correction for publication bias. PET-PEESE was preferred over other methods to correct for bias for three reasons. First, PET-PEESE has been shown to be one of the best performing methods in simulation studies (e.g., Carter et al., 2019; Hong & Reed, 2021). Second, the simulations have favorable conditions for PET-PEESE with 36 and 236 studies included in the meta-analyses and substantial variation in the within-study sampling variances (Niemeyer et al., 2020; Stanley, 2017). Third, PET-PEESE does, in contrast to selection model approaches (e.g., Hedges & Vevea, 2005; Terrin, Schmid, Lau, & Olkin, 2003) not suffer from convergence problems that are difficult to deal with in simulation studies.

The outcome variables in the simulations were bias and root mean squared error (RMSE) for both  $\mu$  and  $\tau^2$ . Another outcome variable was the Type-I error rate/statistical

power for testing the null-hypothesis of no effect. The number of simulated meta-analyses per condition was 10,000. The random-effects model was fitted using the R package “metafor” (Viechtbauer, 2010) using the maximum likelihood estimator for estimating  $\tau^2$ . The maximum likelihood estimator was used to allow for a direct comparison with HYEMA that also uses maximum likelihood estimation. R code of the simulation study is available at <https://osf.io/qwjxn> for the red-romance hypothesis and at <https://osf.io/p843v> for money priming.

## Results

Table 2 shows the results of the simulation studies for the outcome variables bias and RMSE. Values in bold indicate for which method the absolute value of the bias or RMSE was the smallest. The first six columns show the results for estimating  $\mu$ . The main conclusion of the simulation studies with respect to estimating  $\mu$  is that the bias and RMSE was the smallest for the random-effects model in the absence of bias. In the absence of publication bias, HYEMA was approximately unbiased, but it was less efficient than the random-effects model since HYEMA had a larger RMSE. Estimation of the random-effects model was more biased and had larger RMSE than HYEMA when publication bias was present (i.e.,  $pub > 0$ ). PET-PEESE its bias and RMSE were generally larger than HYEMA, and the performance of HYEMA was comparable when PET-PEESE yielded a smaller bias or RMSE.

The last four columns in Table 2 show the results for estimating  $\tau^2$ . A similar pattern was observed for estimating  $\tau^2$  when compared to estimating  $\mu$  although the bias of the random-effects model was less severe and it was the least biased method when publication bias was present in two conditions.

Table 2

Results of the simulation studies to examine the parameter estimation of random-effects meta-analysis (RE MA), PET-PEESE, and HYEMA.

$\mu$	$\tau^2$	$pub$	Bias $\mu$			RMSE $\mu$			Bias $\tau^2$		RMSE $\tau^2$	
			RE MA	PET-PEESE	HYEMA	RE MA	PET-PEESE	HYEMA	RE MA	HYEMA	RE MA	HYEMA
Red-romance hypothesis												
0.128	0.044	0	<b>0.000</b>	-0.030	-0.001	<b>0.053</b>	0.143	0.068	<b>-0.003</b>	-0.003	<b>0.022</b>	0.024
		0.5	0.049	0.020	<b>0.001</b>	0.074	0.145	<b>0.068</b>	0.007	<b>-0.002</b>	0.025	<b>0.023</b>
		0.9	0.204	0.099	<b>0.004</b>	0.211	0.177	<b>0.076</b>	0.017	<b>-0.002</b>	0.033	<b>0.022</b>
		1	0.359	0.027	<b>0.010</b>	0.360	0.154	<b>0.100</b>	<b>-0.002</b>	-0.003	0.025	<b>0.024</b>
0.018	0.009	0	<b>-0.001</b>	-0.012	0.002	<b>0.040</b>	0.096	0.045	<b>-0.000</b>	0.001	<b>0.010</b>	0.012
		0.5	0.017	<b>-0.001</b>	0.002	<b>0.045</b>	0.102	0.045	0.004	<b>0.001</b>	0.013	<b>0.011</b>
		0.9	0.110	0.046	<b>0.001</b>	0.120	0.125	<b>0.049</b>	0.026	<b>-0.000</b>	0.032	<b>0.010</b>
		1	0.386	-0.018	<b>0.000</b>	0.387	<b>0.067</b>	0.074	0.025	<b>-0.000</b>	0.031	<b>0.010</b>
Money priming												
0.319	0.118	0	<b>-0.000</b>	-0.004	0.000	<b>0.026</b>	0.083	0.037	<b>-0.001</b>	-0.001	<b>0.015</b>	0.015
		0.5	0.083	0.044	<b>0.000</b>	0.087	0.092	<b>0.037</b>	-0.001	<b>-0.001</b>	0.015	<b>0.015</b>
		0.9	0.229	0.083	<b>0.001</b>	0.230	0.112	<b>0.042</b>	-0.026	<b>-0.001</b>	0.030	<b>0.017</b>
		1	0.293	0.086	<b>0.002</b>	0.294	0.114	<b>0.048</b>	-0.046	<b>-0.001</b>	0.048	<b>0.018</b>
0.176	0.080	0	<b>-0.000</b>	-0.010	0.000	<b>0.023</b>	0.080	0.030	-0.001	<b>-0.001</b>	<b>0.011</b>	0.012
		0.5	0.068	0.041	<b>0.000</b>	0.072	0.088	<b>0.031</b>	0.008	<b>-0.001</b>	0.014	<b>0.011</b>
		0.9	0.231	0.098	<b>0.001</b>	0.232	0.126	<b>0.036</b>	<b>-0.000</b>	-0.001	0.012	<b>0.011</b>
		1	0.335	0.087	<b>0.002</b>	0.335	0.132	<b>0.043</b>	-0.023	<b>-0.001</b>	0.025	<b>0.013</b>

Note: PET-PEESE is based on weighted least squares regression and does not estimate  $\tau^2$ . RMSE = Root Mean Squared Error

Table 3 shows the Type-I error rate for the test of no effect with bold values indicating the method that yielded a Type-I error rate closest to 0.05. Conducting the hypothesis test failed in at most 3% of the simulated meta-analyses for HYEMA due to numerical problems with inverting the Hessian (see Table S1 available at <https://osf.io/jq6dc> for detailed results per condition). The results in Table 3 are based on all simulated meta-analyses for which the hypothesis did not fail for a particular method. The Type-I error rate of the random-effects model was close to 0.05 in the absence of publication bias. If publication bias was present, the Type-I error rate of the random-effects model was severely inflated. PET-PEESE its Type-I error rate was too high except for one condition in the simulations based on the meta-analysis on the red-romance hypothesis where it was too low. The Type-I error rate of HYEMA was close to 0.05 for all conditions and its Type-I error rate was closer to 0.05 when compared to the other methods for the vast majority of conditions.

The results on statistical power are only presented in the supplemental materials (Table S2 available at <https://osf.io/jq6dc>), because these results of random-effects model and PET-PEESE are distorted by the generally severely inflated Type-I error rate. Statistical power of HYEMA for the red-romance hypothesis was approximately 0.4 for the conditions with  $\mu=0.128$  and close to the Type-I error rate for the conditions with  $\mu=0.018$ . For the money priming meta-analysis, HYEMA's statistical power was close to 1 for all conditions.

## Discussion

This paper has introduced HYEMA that takes into account whether a study is a conventional or preregistered study in a meta-analysis. This may yield improved meta-analytic conclusions compared to the random-effects model since there is corrected for publication bias in the conventional studies. HYEMA was applied to two meta-analyses on prominent effects in the psychological literature, and the results showed that the average effect size estimates were substantially reduced and closer to zero for both meta-analyses. Two simulation studies tailored towards the two meta-analyses showed that HYEMA had

Table 3

*Results of the simulation studies to examine the Type-I error rate for the test of no effect of random-effects meta-analysis (RE MA), PET-PEESE, and HYEMA.*

$\mu$	$\tau^2$	<i>pub</i>	RE MA	PET-PEESE	HYEMA
Red-romance hypothesis					
0	0.044	0.0	0.069	0.140	<b>0.066</b>
		0.5	0.114	0.175	<b>0.063</b>
		0.9	0.812	0.384	<b>0.061</b>
		1.0	1.000	0.143	<b>0.070</b>
0	0.009	0.0	0.063	0.083	<b>0.052</b>
		0.5	0.082	0.094	<b>0.058</b>
		0.9	0.531	0.178	<b>0.054</b>
		1.0	1.000	0.012	<b>0.041</b>
Money priming					
0	0.118	0.0	<b>0.052</b>	0.366	0.052
		0.5	0.656	0.487	<b>0.051</b>
		0.9	1.000	0.649	<b>0.052</b>
		1.0	1.000	0.530	<b>0.056</b>
0	0.080	0.0	<b>0.049</b>	0.347	0.048
		0.5	0.554	0.464	<b>0.054</b>
		0.9	1.000	0.649	<b>0.051</b>
		1.0	1.000	0.458	<b>0.054</b>

*Note:* Bold values indicate the method with the Type-I error rate that deviated the least from  $\alpha = 0.05$ .

smaller bias and lower RMSE for estimating the average effect size than the random-effects model if publication bias was present. Furthermore, HYEMA's bias was also smaller and its RMSE lower compared to PET-PEESE for the majority of conditions. Hence, HYEMA is recommended to be applied as a sensitivity analysis in meta-analyses that contain a mix of conventional and preregistered studies.

HYEMA was introduced by assuming that the goal of the meta-analysis was to estimate and provide statistical inference for the average effect size and between-study variance in true effect sizes. Another goal of the meta-analysis is often to explain the between-study variance by including moderators (Higgins, Thompson, & Spiegelhalter, 2009). This goal can be achieved with HYEMA by estimating a linear regression equation that is a

function of the moderators of interest rather than only estimating the average true effect size. This approach for incorporating moderators in HYEMA is akin to how this is done for the random-effects model (e.g., Van Houwelingen, Arends, & Stijnen, 2002). The supplemental materials available at <https://osf.io/f7g4x> show how the analyses can be conducted if a moderator variable is included in the example meta-analysis on the red-romance hypothesis.

It is assumed that the average true effect size and the between-study variance is the same for all conventional and preregistered studies in HYEMA. Whether this assumption holds in practice is an empirical question that can be examined by using HYEMA. A dichotomous moderator indicating whether a study is a conventional or preregistered study can be included in HYEMA to study whether there are differences in average true effect size. HYEMA can also be fitted once to the data of only the conventional studies and once to the data of only the preregistered studies to assess whether the between-study variance is different for the two sets of studies.

Only preregistered studies were treated as non-conventional studies in the applications. However, Nosek et al. (2022) argued that meta-analyzing conventional and replication studies can yield a false sense of precision and accuracy if the studies differ in quality and risk of bias. A difference in risk of bias can be expected, because replications are probably more likely to be published irrespective of their results. This calls for treating both preregistered *and* replication studies as being different from conventional studies, and this is an additional sensitivity analysis that can be conducted using HYEMA. This analysis was conducted for the meta-analysis on the red-romance hypothesis by first coding whether each conventional study was a replication. Nine of the 28 (32.1%) conventional studies were replications<sup>3</sup>. The

---

<sup>3</sup> A study was coded as a replication if it was explicitly stated in the paper that another study was replicated. Nine replications was probably an underestimate of the total number of replications, because there was not enough information for eight effect sizes to code whether the study was a replication. Effect sizes for these eight studies were, for example, computed based on unpublished data and no other information was available.



percentage of statistically significant conventional studies increased from 35.7% to 47.4% when the replications were no longer considered to be conventional studies. This is also indicative that the replications were different from the conventional studies. Applying HYEMA under the assumption that preregistered and replication studies are different from conventional studies yielded comparable results when comparing this to the initial application of the method. Estimates of the average effect size and between-study variance in true effect sizes were now slightly closer to zero ( $\hat{\mu}$ =0.009 vs. 0.018;  $\hat{\tau}^2$ =0.006 vs. 0.009).

A limitation of this paper is that the simulation studies were tailored to characteristics of the two applications. Future research may focus on studying the statistical properties of HYEMA in simulated data that are representative for psychological research in general. An interesting question that warrants further study in simulations is whether there are characteristics of a meta-analysis where HYEMA should become the main analysis rather than a sensitivity analysis. HYEMA should probably especially be favored over the random-effects model if publication bias is severe and the proportion of preregistered studies in the meta-analysis is small. Overestimation of the average effect size of the random-effects model is then large. Conclusions regarding this cannot be drawn based on the simulation studies in this paper, because the number of preregistered studies was not systematically varied.

A limitation of HYEMA is that  $p$ -hacking in the included studies in the meta-analysis likely biases the parameter estimation and distorts the statistical inference. The detrimental effect of  $p$ -hacking is to be expected, because closely related methods such as  $p$ -uniform and the three-parameter selection model approach are known to not perform well if  $p$ -hacking is present (Carter et al., 2019; van Aert et al., 2016). However, it has to be emphasized that none of the current meta-analysis methods perform well if the included studies are  $p$ -hacked.

$P$ -hacking could be utilized in the conventional studies to get desirable results but also in preregistered studies whereas HYEMA assumes that the preregistered studies are unbiased.

*P*-hacking in the preregistered studies is more likely to be utilized if the preregistration is of low quality and not specific enough. Research has shown that preregistrations often lack details such that there is still leeway for researchers to analyze their data in multiple ways and that researchers sometimes do not disclose deviations from the preregistration (Claesen, Gomes, Tuerlinckx, & Vanpaemel, 2021; van den Akker et al., 2023). There can also be deviations from the preregistration by using so-called reverse *p*-hacking to turn an initially statistically significant result in a nonsignificant result (Chuard, Vrtílek, Head, & Jennions, 2019; Protzko, 2018). A nonsignificant result can be desirable to debunk a prominent study. If there are doubts about the quality of preregistered studies, a sensitivity analysis can be conducted using HYEMA where the preregistered studies that are deemed to be of low quality are treated as conventional studies or are excluded from the analysis. Another option is to make use of the risk of bias assessment that is recommended to be done in every meta-analysis (e.g., Appelbaum et al., 2018; Valentine, 2019). Such a risk of bias assessment often contains items on whether participants were randomly assigned to conditions, whether covariates are included in the model, and also often whether a study was preregistered or not. The scores of the risk of bias assessment could be included as a moderator in HYEMA to examine whether there is a relationship between the study's effect size and its risk of bias.

Future research could be devoted to assess to what extent HYEMA is affected by *p*-hacking. It could also be studied whether the proposed approach of van Aert and Wicherts (2023) can be incorporated in HYEMA to correct for selective outcome reporting, which is the most common *p*-hacking behavior (John, Loewenstein, & Prelec, 2012). Future research could also focus on increasing the modelling flexibility of HYEMA. The computation of the likelihood currently depends on whether a study is statistically significant or not, but there is not differentiated in how the likelihood is computed within the set of significant and nonsignificant studies. For example, statistically nonsignificant studies with a positive effect size may have a larger probability of getting published than nonsignificant studies with a negative effect size (Hedges & Vevea, 1996). More differentiation in how the likelihood is

492 computed might yield less biased estimators and better statistical inference.

493       HYEMA can be applied using the `hybrid()` function in the R package “puniform”  
494 (van Aert, 2023). The only extra information that is required for applying this function  
495 compared to fitting the random-effects model is an argument specifying which studies are  
496 conventional studies. A step-by-step video tutorial where applying HYEMA is illustrated is  
497 available at <https://osf.io/b9m5v>. HYEMA can also be applied by using the web application  
498 <https://rcmvanaert.shinyapps.io/HYEMA> that does not require any knowledge about  
499 programming in R.

500       To conclude, HYEMA is the first method that treats conventional and preregistered  
501 studies differently in a meta-analysis and only corrects for publication bias in the  
502 conventional studies. Applying HYEMA may yield less biased estimates and better  
503 statistical inference if publication bias is present in the conventional studies. Hence,  
504 meta-analysts are recommended to apply the method as a sensitivity analysis if both  
505 conventional and preregistered studies are included in their meta-analysis.

### Conflicts of Interest

The author declares that there were no conflicts of interest with respect to the authorship or the publication of this article.

### Acknowledgements

I would like to thank the members of the Meta-Research Center at Tilburg University for providing feedback on an earlier version of this paper.

### Funding

Robbie C.M. van Aert is supported by a Veni grant financed by the Dutch Research Council (NWO). Grant Number: VI.Veni.211G.012.

### Prior versions

A preprint of this manuscript is available at <https://osf.io/preprints/metaarxiv/2bj85/>.

### Data availability statement

The data of the two applications as well as R code that was used for the analyses and simulation studies is available via the OSF project page (<https://osf.io/qbexm/>).

## References

- Agresti, A. (2013). *Categorical data analysis*. Hoboken, N.J.: Wiley-Interscience.
- Andrews, D. W. K. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, 69(3), 683–734. Retrieved from <https://www.jstor.org/stable/2692206>
- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *The American Psychologist*, 73(1), 3–25. <https://doi.org/10.1037/amp0000191>
- Bartoš, F., Maier, M., Shanks, D. R., Stanley, T. D., Sladekova, M., & Wagenmakers, E.-J. (2023). Meta-analyses in psychology often overestimate evidence for and size of effects. *Royal Society Open Science*, 10(7), 230224. <https://doi.org/10.1098/rsos.230224>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: John Wiley & Sons, Ltd.
- Camerer, C. F., Dreber, A., Forsell, E., Ho, T. H., Huber, J., Johannesson, M., ... Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280), 1433–1436. <https://doi.org/10.1126/science.aaf0918>
- Carter, E. C., & McCullough, M. E. (2018). A simple, principled approach to combining evidence from meta-analysis and high-quality replications. *Advances in Methods and Practices in Psychological Science*, 1(2), 174–185. <https://doi.org/10.1177/2515245918756858>
- Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science*, 2(2), 115–144. <https://doi.org/10.1177/2515245919847196>
- Casella, G., & Berger, R. L. (2002). *Statistical Inference*. Belmont, CA: Duxbury.

- Chambers, C. D., & Munafò, M. R. (2013). Trust in science would be improved by study pre-registration. <https://www.theguardian.com/science/blog/2013/jun/05/trust-in-science-study-pre-registration>.
- Chuard, P. J. C., Vrtílek, M., Head, M. L., & Jennions, M. D. (2019). Evidence that nonsignificant results are sometimes preferred: Reverse p-hacking or selective reporting? *PLOS Biology*, 17(1), e3000127. <https://doi.org/10.1371/journal.pbio.3000127>
- Claesen, A., Gomes, S., Tuerlinckx, F., & Vanpaemel, W. (2021). Comparing dream to reality: An assessment of adherence of the first generation of preregistered studies. *Royal Society Open Science*, 8(10), 211037. <https://doi.org/10.1098/rsos.211037>
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315(7109), 629–634. <https://doi.org/10.1136/bmj.315.7109.629>
- Ensinnck, E. N. F., & Lakens, D. (2023). *An inception cohort study quantifying how many registered studies are published* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/5hkjz>
- Errington, T. M., Mathur, M., Soderberg, C. K., Denis, A., Perfito, N., Iorns, E., & Nosek, B. A. (2021). Investigating the replicability of preclinical cancer biology. *eLife*, 10, e71601. <https://doi.org/10.7554/eLife.71601>
- Francis, G. (2013). Publication bias in “Red, rank, and romance in women viewing men,” by Elliot et al. (2010). *Journal of Experimental Psychology: General*, 142(1), 292–296. <https://doi.org/10.1037/a0027923>
- Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203), 1502–1505. <https://doi.org/10.1126/science.1255484>

- Hardy, R. J., & Thompson, S. G. (1996). A likelihood approach to meta-analysis with random effects. *Statistics in Medicine*, 15(6), 619–629.  
[https://doi.org/10.1002/\(SICI\)1097-0258\(19960330\)15:6%3C619::AID-SIM188%3E3.0.CO;2-A](https://doi.org/10.1002/(SICI)1097-0258(19960330)15:6%3C619::AID-SIM188%3E3.0.CO;2-A)
- Hedges, L. V. (1984). Estimation of effect size under nonrandom sampling: The effects of censoring studies yielding statistically insignificant mean differences. *Journal of Educational Statistics*, 9(1), 61–85.
- Hedges, L. V., & Vevea, J. L. (1996). Estimating effect size under publication bias: Small sample properties and robustness of a random effects selection model. *Journal of Educational and Behavioral Statistics*, 21(4), 299–332.  
<https://doi.org/10.2307/1165338>
- Hedges, L. V., & Vevea, J. L. (2005). Selection method approaches. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 145–174). Chichester: UK: Wiley.
- Higgins, J. P. T., Thompson, S. G., & Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society*, 172(1), 137–159. <https://doi.org/10.1111/j.1467-985X.2008.00552.x>
- Hong, S., & Reed, W. R. (2021). Using Monte Carlo experiments to select meta-analytic estimators. *Research Synthesis Methods*, 12(2), 192–215.  
<https://doi.org/10.1002/jrsm.1467>
- Iyengar, S., & Greenhouse, J. B. (1988). Selection models and the file drawer problem: Rejoinder. *Statistical Science*, 3(1), 133–135.
- Jackson, D., & White, I. R. (2018). When should meta-analysis avoid making hidden normality assumptions? *Biometrical Journal*, 60(6), 1040–1058.  
<https://doi.org/10.1002/bimj.201800071>
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological*

601 *Science*, 23(5), 524–532. <https://doi.org/10.1177/0956797611430953>

602 Konstantopoulos, S., & Hedges, L. V. (2019). Statistically analyzing effect sizes:  
603 Fixed- and random-effects models. In H. Cooper, L. V. Hedges, & J. C. Valentine  
604 (Eds.), *The handbook of research synthesis and meta-analysis* (3rd ed., pp.  
605 245–280). New York: Russell Sage Foundation.

606 Lehmann, G. K., Elliot, A. J., & Calin-Jageman, R. J. (2018). Meta-analysis of the  
607 effect of red on perceived attractiveness. *Evolutionary Psychology*, 16(4),  
608 147470491880241. <https://doi.org/10.1177/1474704918802412>

609 Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing*  
610 *research*. Cambridge, MA: Harvard University Press.

611 Lodder, P., Ong, H. H., Grasman, R. P. P. P., & Wicherts, J. M. (2019). A  
612 comprehensive meta-analysis of money priming. *Journal of Experimental*  
613 *Psychology: General*, 148(4), 688–712. <https://doi.org/10.1037/xge0000570>

614 Makel, M. C., Hodges, J., Cook, B. G., & Plucker, J. A. (2021). Both questionable  
615 and open research practices are prevalent in education research. *Educational*  
616 *Researcher*, 50(8), 493–504. <https://doi.org/10.3102/0013189X211001356>

617 McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication  
618 bias in meta-analysis: An evaluation of selection methods and some cautionary  
619 notes. *Perspectives on Psychological Science*, 11(5), 730–749.  
620 <https://doi.org/10.1177/1745691616662243>

621 Molenberghs, G., & Verbeke, G. (2007). Likelihood ratio, score, and Wald tests in a  
622 constrained parameter space. *The American Statistician*, 61(1), 22–27.  
623 <https://doi.org/10.1198/000313007X171322>

624 Niemeyer, H., van Aert, R. C. M., Schmid, S., Uelsmann, D., Knaevelsrud, C., &  
625 Schulte-Herbrueggen, O. (2020). Publication bias in meta-analyses of  
626 posttraumatic stress disorder interventions. *Meta-Psychology*, 4.  
627 <https://doi.org/10.15626/MP.2018.884>



- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600–2606. <https://doi.org/10.1073/pnas.1708274114>
- Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., . . . Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, 73(1), 719–748. <https://doi.org/10.1146/annurev-psych-020821-114157>
- Nosek, B. A., & Lindsay, D. S. (2018). Preregistration becoming the norm in psychological science. *APS Observer*, 31.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251). <https://doi.org/10.1126/science.aac4716>
- Pawitan, Y. (2013). *In all likelihood: Statistical modelling and inference using likelihood*. Oxford: OUP Oxford.
- Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, 61(10), 991–996. <https://doi.org/10.1016/j.jclinepi.2007.11.010>
- Protzko, J. (2018). *Null-hacking, a lurking problem* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/9y3mp>
- R Core Team. (2023). *R: A language and environment for statistical computing*.
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). Publication bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Chichester, UK: Wiley.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366.

655 <https://doi.org/10.1177/0956797611417632>

656 Simonsohn, U. (2023). Preregistration prevalence. Retrieved November 20, 2023,  
657 from <https://datacolada.org/115>.

658 Stanley, T. D. (2017). Limitations of PET-PEESE and other meta-analysis methods.  
659 *Social Psychological and Personality Science*, 8(5), 581–591.

660 <https://doi.org/10.1177/1948550617693062>

661 Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce  
662 publication selection bias. *Research Synthesis Methods*, 5(1), 60–78.

663 <https://doi.org/10.1002/jrsm.1095>

664 Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. (2003). Adjusting for publication bias  
665 in the presence of heterogeneity. *Statistics in Medicine*, 22(13), 2113–2126.

666 <https://doi.org/10.1002/sim.1461>

667 Vadillo, M. A., Hardwicke, T. E., & Shanks, D. R. (2016). Selection bias, vote  
668 counting, and money-priming effects: A comment on Rohrer, Pashler, and Harris  
669 (2015) and Vohs (2015). *Journal of Experimental Psychology: General*, 145(5),  
670 655–663. <https://doi.org/10.1037/xge0000157>

671 Valentine, J. C. (2019). Incorporating judgments about study-quality into research  
672 synthesis. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of*  
673 *research synthesis and meta-analysis* (3rd ed., pp. 129–140). New York: Russell  
674 Sage Foundation.

675 van Aert, R. C. M. (2023). *Puniform: Meta-analysis methods correcting for*  
676 *publication bias*.

677 van Aert, R. C. M., & van Assen, M. A. L. M. (2018). Examining reproducibility in  
678 psychology: A hybrid method for combining a statistically significant original  
679 study and a replication. *Behavior Research Methods*, 50(4), 1515–1539.

680 <https://doi.org/10.3758/s13428-017-0967-6>

681 van Aert, R. C. M., & van Assen, M. A. L. M. (2024). *Correcting for publication bias*

in a meta-analysis with the *p*-uniform\* method. Manuscript submitted for publication. <https://doi.org/10.31222/osf.io/zqjr9>

van Aert, R. C. M., & Wicherts, J. M. (2023). Correcting for outcome reporting bias in a meta-analysis: A meta-regression approach. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-023-02132-2>

van Aert, R. C. M., Wicherts, J. M., & van Assen, M. A. L. M. (2016). Conducting meta-analyses on *p*-values: Reservations and recommendations for applying *p*-uniform and *p*-curve. *Perspectives on Psychological Science*, 11(5), 713–729. <https://doi.org/10.1177/1745691616650874>

van Assen, M. A. L. M., van Aert, R. C. M., & Wicherts, J. M. (2015). Meta-analysis using effect size distributions of only statistically significant studies. *Psychological Methods*, 20(3), 293–309. <https://doi.org/10.1037/met0000025>

van den Akker, O. R., Bakker, M., van Assen, M. A. L. M., Pennington, C. R., Verweij, L., Elsherif, M. M., . . . Wicherts, J. M. (2023). *The effectiveness of preregistration in psychology: Assessing preregistration strictness and preregistration-study consistency* [Preprint]. MetaArXiv. <https://doi.org/10.31222/osf.io/h8xjw>

van den Akker, O. R., Scherer, L. D., Wicherts, J. M., & Koole, S. (2020). *Support for open science practices in emotion science: A survey study*. <https://doi.org/10.31234/osf.io/ub4wc>.

Van Houwelingen, H. C., Arends, L. R., & Stijnen, T. (2002). Advanced methods in meta-analysis: Multivariate approach and meta-regression. *Statistics in Medicine*, 21(4), 589–624. <https://doi.org/10.1002/sim.1040>

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, 60(3), 419–435. <https://doi.org/10.1007/bf02294384>

Viechtbauer, W. (2007a). Confidence intervals for the amount of heterogeneity in

meta-analysis. *Statistics in Medicine*, 26(1), 37–52.

<https://doi.org/10.1002/sim.2514>

Viechtbauer, W. (2007b). Hypothesis tests for population heterogeneity in

meta-analysis. *British Journal of Mathematical and Statistical Psychology*, 60,

29–60. <https://doi.org/10.1348/000711005X64042>

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>

Viechtbauer, W. (2021). Model checking in meta-analysis. In C. H. Schmid, T.

Stijnen, & I. R. White (Eds.), *Handbook of meta-analysis* (1st ed.). Abingdon:

Chapman and Hall.

Vohs, K. D., Mead, N. L., & Goode, M. R. (2006). The psychological consequences of

money. *Science*, 314(5802), 1154–1156. <https://doi.org/10.1126/science.1132491>

Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A.

(2012). An agenda for purely confirmatory research. *Perspectives on Psychological*

*Science*, 7(6), 632–638. <https://doi.org/10.1177/1745691612463078>

Washburn, A. N., Hanson, B. E., Motyl, M., Skitka, L. J., Yantis, C., Wong, K. M.,

... Carsel, T. S. (2018). Why do some psychology researchers resist adopting

proposed reforms to research practices? A description of researchers' rationales.

*Advances in Methods and Practices in Psychological Science*, 1(2), 166–173.

<https://doi.org/10.1177/2515245918757427>

White, T., Noble, D., Senior, A., Hamilton, W., & Viechtbauer, W. (2022). *Metadat:*

*Meta-Analysis Datasets*.

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R.

C. M., & van Assen, M. A. L. M. (2016). Degrees of freedom in planning, running,

analyzing, and reporting psychological studies: A checklist to avoid p-hacking.

*Frontiers in Psychology*, 7(1832). <https://doi.org/10.3389/fpsyg.2016.01832>