

DATA PAPER

Estimates of Between-Study Heterogeneity for 705 Meta-Analyses Reported in *Psychological Bulletin* From 1990–2013

Sara van Erp¹, Josine Verhagen², Raoul P. P. Grasman² and Eric-Jan Wagenmakers²¹ Tilburg University, NL² University of Amsterdam, NL

Corresponding author: Sara van Erp (s.vanerp_1@tilburguniversity.edu)

We present a data set containing 705 between-study heterogeneity estimates τ^2 as reported in 61 articles published in *Psychological Bulletin* from 1990–2013. The data set also includes information about the number and type of effect sizes, the Q - and I^2 -statistics, and publication bias. The data set is stored in the Open Science Framework repository (<https://osf.io/wyhve/>) and can be used for several purposes: (1) to compare a specific heterogeneity estimate to the distribution of between-study heterogeneity estimates in psychology; (2) to construct an informed prior distribution for the between-study heterogeneity in psychology; (3) to obtain realistic population values for Monte Carlo simulations investigating the performance of meta-analytic methods.

Keywords: heterogeneity; meta-analysis; Bayesian inference; informed prior distribution**Funding statement:** This research was supported by the ERC project “Bayes or Bust”.

Overview

Context

Collection date

The data were collected between July 2014 and September 2014.

Background

Meta-analysis provides an important tool to synthesize evidence across different studies on the same topic. For the analysis the researcher typically needs to adopt either a fixed-effect model or a random-effects model. A fixed-effect meta-analysis assumes one true underlying effect size with variation in observed effect sizes arising solely because of sampling error. In contrast, a random-effects meta-analysis assumes a distribution of true effect sizes; consequently, observed differences in effect sizes arise both from sampling error and from true differences between effect sizes. The latter form of variation, or between-study heterogeneity, is the focus of this data set.

The random-effects model for $i = 1, \dots, n$ independent effect size estimates y from n studies is given by:

$$y_i = \theta_i + \epsilon_i, \quad (1)$$

where the sampling error ϵ_i is assumed to be normally distributed, i.e., $\epsilon_i \sim N(0, \sigma_\epsilon^2)$, and the effect sizes θ_i are assumed to follow a normal distribution with mean μ and variance

τ^2 , i.e., $\theta_i \sim N(\mu, \tau^2)$. The estimate of the between-study heterogeneity, or τ^2 , is an interesting parameter since it provides information about the robustness of the effect across different contexts. If the heterogeneity is large, the effect shows considerable variation across studies, and the researcher might consider to investigate possible moderators. For example, the meta-analysis by Snyder (2013) on cognitive impairments in patients with major depressive disorder found that effect sizes for verbal working memory were larger for patients receiving a specific type of medication, compared to patients who did not receive this medication.

It is unclear, however, what constitutes a “large” heterogeneity in the field of psychology. Therefore, it is important to obtain an overview of between-study heterogeneity estimates encountered in meta-analyses in psychology. Such an overview allows researchers to compare the τ^2 estimate in their meta-analysis to the general distribution of estimates and gauge the size of their between-study heterogeneity. In addition, when conducting a Bayesian meta-analysis, this overview can provide a basis for the construction of an informed prior distribution for between-study heterogeneity.

In the medical literature, several attempts have been made to present an overview of between-study heterogeneity estimates from a large number of meta-analyses and construct an informed prior distribution

based on these estimates [1, 2, 3]. In the field of psychology, an early overview of meta-analyses on the efficacy of psychological, educational, and behavioral treatments is given by [4]. Based on the resulting 302 meta-analyses, [4] investigated several characteristics, including effect size, methodological quality, publication bias, and sample size. A second overview in psychology is given by [5], who focused on the use of fixed-effect and random-effects models in meta-analyses reported in *Psychological Bulletin* from 1978 to 2006 (169 studies). However, none of the existing overviews of meta-analyses in psychology include between-study heterogeneity. Therefore, the aim of this study was to provide a general distribution of between-study heterogeneity estimates, based on τ^2 estimates extracted from a large number of meta-analyses in the field of psychology. The data set presented here is the result of this literature review.

Methods

Sample

We extracted 705 between-study heterogeneity estimates τ^2 from meta-analyses reported in 61 articles published in *Psychological Bulletin* from 1990–2013. Most articles featured different outcome measures and therefore provided multiple estimates. For example, when considering sex differences in sexuality, Petersen and Hyde (2010) considered 14 sexual behaviors and 16 sexual attitudes, resulting in 30 heterogeneity estimates. Therefore, the heterogeneity estimates are not fully independent of each other.

The selected time frame featured 255 articles reporting meta-analyses, but we included only studies that provided estimates of the between-study variance τ^2 or standard deviation τ . Some studies performed the meta-analysis with multiple models (e.g., fixed and random effects, with and without moderators). In these cases, we included only the random effects analysis without moderators, in order to obtain a maximum indication of the between-study heterogeneity.

Materials

No materials were used, apart from a laptop to search for meta-analyses.

Procedures

First, we searched for articles containing overviews of meta-analyses; however, these articles almost never reported the estimated τ^2 values. Instead, we focused on separate meta-analyses published in *Psychological Bulletin*, a journal that primarily publishes qualitative and quantitative reviews in psychology, as is evident from the journal description:

“*Psychological Bulletin* publishes evaluative and integrative research reviews and interpretations of issues in scientific psychology. Both qualitative (narrative) and quantitative (meta-analytic) reviews will be considered, depending on the nature of the database under consideration for review.” [6].

Through the search engine Ebscohost, we searched each issue from 1990 up to and including 2013 by

selecting “meta-analysis” as methodology or by searching on the keyword “meta-analysis”. If this gave no results, all titles in the issue were scanned for the keywords “meta-analysis”, “review”, or “synthesis”. The resulting 255 articles were scanned to determine whether a meta-analysis was performed and an estimate of τ^2 provided. This was the case in 61 articles. Since most articles included multiple analyses (such as separate meta-analyses for multiple outcome variables), we extracted several heterogeneity estimates per article, resulting in a final data set of 705 estimates. Note that different estimators for τ^2 exist (e.g., Hunter-Schmidt, DerSimonian-Laird, maximum likelihood), which can provide varying or even conflicting estimates of the between-study heterogeneity [7]. Unfortunately, most articles did not report which estimator for τ^2 was used. All meta-analyses were based on the random-effects model in equation 1. When articles provided the between-study standard deviation (variance), we computed the variance (standard deviation) so that the data set includes both τ and τ^2 estimates for each meta-analysis. The data extraction and coding were performed manually by the first author.

Quality control

The data were collected with the utmost care and conscientiousness. In addition, the Appendix provides an overview of all articles included in the data set so that the full data set can be easily checked.

Ethical issues

Only secondary data were used to construct the data set.

Data set description

Figure 1 shows a screenshot of the data file. For each article, the following information was recorded: (1) the variables in the meta-analysis; (2) the number of effect sizes per analysis; (3) the between-study standard deviation τ ; (4) the between-study variance τ^2 ; (5) the Q -statistic (i.e., the test statistic of a commonly used statistical test to determine whether there exists significant heterogeneity; [8]); (6) the I^2 -statistic (i.e., the percentage of total variation in effect sizes due to true between-study heterogeneity; computed based on the Q -statistic as: $\frac{Q-df}{Q} * 100$; [9]); (7) the type of effect size; (8) the test(s) used to assess publication bias; and (9) whether or not publication bias was present. We have included a recoded variable for publication bias with ‘No publication bias’ = 0 and ‘Publication bias’ = 1. Publication bias is coded as ‘NA’ if no test for publication bias was reported in the article. This was the case in 385 instances. Histograms of the between-study standard deviation for mean differences and correlations are available in **Figure 2** and **Figure 3**, respectively. Note that 18 heterogeneity estimates were not based on Cohen’s d , Hedges’ g , or Pearson r effect sizes and these estimates have not been included in the histograms. **Figure 4** shows the histogram for the I^2 -statistic, which does not depend on the type of effect size used. Note that the I^2 statistic is fixed to zero when Q is smaller than the degrees of freedom (i.e., the number of effect sizes – 1).

#	A	B	C	D	E	F	G	H	I	J	K	L
1	Reference	Entry in reference description	# of effect sizes	tau	tau ²	Q	I ²	Type of ES	Test for publication bias	Publication bias	Publication bias recoded	
2	Snyder 2013	1 inhibition	48	0,3	0,09	109,67	57,1442	Cohen's d	trim-and-fill	no	no	0
3	Snyder 2013	2 shifting	69	0,244949	0,06	133,83	49,1893	Cohen's d	trim-and-fill	no	no	0
4	Snyder 2013	3 updating	10	0,509902	0,26	35,22	74,4463	Cohen's d	trim-and-fill	no	no	0
5	Snyder 2013	4 verbal working memory	39	0,2	0,04	65,59	42,0643	Cohen's d	trim-and-fill	no	no	0
6	Snyder 2013	5 visuospatial working memory	23	0,2236068	0,05	38,01	42,1205	Cohen's d	trim-and-fill	no	no	0
7	Snyder 2013	6 planning	17	0,1732051	0,03	23,4	31,6239	Cohen's d	trim-and-fill	no	no	0
8	Snyder 2013	7 verbal fluency	46	0,2645751	0,07	100,43	55,1927	Cohen's d	trim-and-fill	no	no	0
9	Snyder 2013	8 psychomotor speed	23	0,3816625	0,11	63,24	65,2119	Cohen's d	trim-and-fill	no	no	0
10	Snyder 2013	9 digit-symbol substitution	22	0,4123106	0,17	88,76	76,3407	Cohen's d	trim-and-fill	no	no	0
11	Snyder 2013	10 vocabulary	36	0,1732051	0,03	47,81	26,7936	Cohen's d	trim-and-fill	no	no	0
12	Uttal et al. 2013	1 training spatial skills	206	0,4801163	0,185			Hedges' g	fail-safe N	no	no	0
13	Sweeny & Krizan 2013	1 temporal decline in expectations	71	0,4	0,16	3339,5	97,9039	Cohen's d	fail-safe N/trim-and-fill	yes	yes	1
14	Balilet & van Lange 2011	1 expectations and cooperation	142	0,3464102	0,12			Pearson's r	trim-and-fill/Egger test	no	no	0
15	Balilet & van Lange 2011	2 trust and cooperation	60	0,1732051	0,03			Pearson's r	trim-and-fill/Egger test	no	no	0
16	Freund & Kasten 2012	1 self-estimates of cognitive ability	154	0,1224745	0,015			Pearson's r	funnel plot	no	no	0
17	Richardson et al. 2012	1 socioeconomic status and GPA	21	0	0	221,26	90,9609	Corrected Pearson's r	trim-and-fill	yes	yes	1
18	Richardson et al. 2012	2 sex and GPA	21	0,01	0,0001	121,9	83,5931	Corrected Pearson's r	trim-and-fill	yes	yes	1
19	Richardson et al. 2012	3 age and GPA	17	0,01	0,0001	353,49	95,4737	Corrected Pearson's r	trim-and-fill	yes	yes	1
20	Richardson et al. 2012	4 high school GPA and GPA	46	0,03	0,0009	1368,25	96,7111	Corrected Pearson's r	trim-and-fill	yes	yes	1
21	Richardson et al. 2012	5 SAT and GPA	29	0,01	0,0001	258,59	89,172	Corrected Pearson's r	trim-and-fill	yes	yes	1
22	Richardson et al. 2012	6 ACT and GPA	21	0,01	0,0001	314,49	93,6405	Corrected Pearson's r	trim-and-fill	yes	yes	1
23	Richardson et al. 2012	7 A level points and GPA	4	0,01	0,0001	12,07	75,145	Corrected Pearson's r	trim-and-fill	yes	yes	1
24	Richardson et al. 2012	8 intelligence and GPA	35	0,01	0,0001	117,94	71,1718	Corrected Pearson's r	trim-and-fill	yes	yes	1
25	Richardson et al. 2012	9 conscientiousness and GPA	69	0	0	165,12	58,8178	Corrected Pearson's r	trim-and-fill	yes	yes	1
26	Richardson et al. 2012	10 procrastination and GPA	10	0	0	13,77	34,6405	Corrected Pearson's r	trim-and-fill	yes	yes	1
27	Richardson et al. 2012	11 openness and GPA	52	0	0	118,6	56,9983	Corrected Pearson's r	trim-and-fill	yes	yes	1
28	Richardson et al. 2012	12 neuroticism and GPA	58	0,01	0,0001	163,7	65,1802	Corrected Pearson's r	trim-and-fill	yes	yes	1
29	Richardson et al. 2012	13 agreeableness and GPA	47	0	0	103,05	55,3615	Corrected Pearson's r	trim-and-fill	yes	yes	1
30	Richardson et al. 2012	14 extraversion and GPA	58	0	0	137,35	58,5002	Corrected Pearson's r	trim-and-fill	yes	yes	1
31	Richardson et al. 2012	15 need for cognition and GPA	5	0,01	0,0001	22,08	81,8841	Corrected Pearson's r	trim-and-fill	yes	yes	1
32	Richardson et al. 2012	16 emotional intelligence and GPA	14	0	0	21,37	39,1671	Corrected Pearson's r	trim-and-fill	yes	yes	1
33	Richardson et al. 2012	17 locus of control and GPA	13	0,02	0,0004	44,85	73,2441	Corrected Pearson's r	trim-and-fill	yes	yes	1

Figure 1: Screenshot of the data file "Data 1990–2013 with tau values.xlsx".

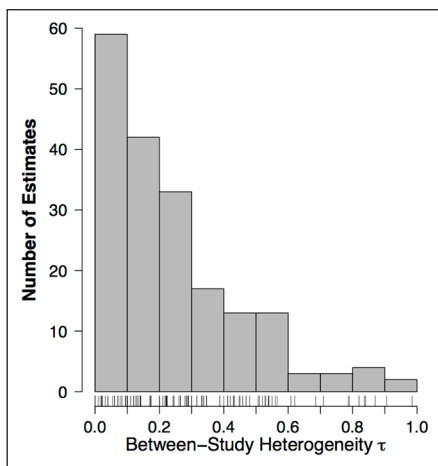


Figure 2: Histogram of the between-study standard deviation for mean difference effect sizes (Cohen's *d* and Hedges' *g*).

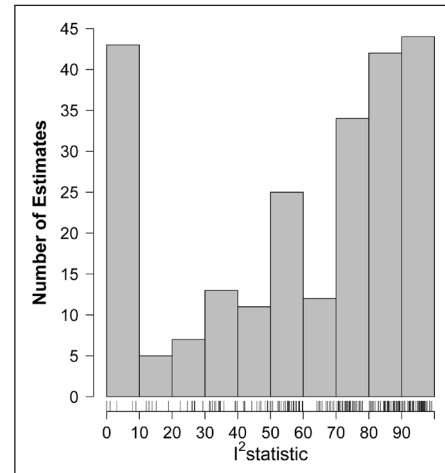


Figure 4: Histogram of the *I*² statistic for all types of effect sizes.

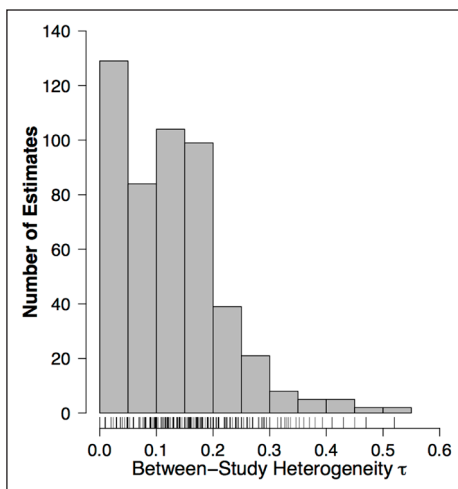


Figure 3: Histogram of the between study standard deviation for correlation effect sizes (Pearson's *r*).

Data type

Secondary data.

Format names and versions

This is the third and final version of the data, in Excel, comma-separated values (CSV), and text format. In the first version, the type of effect size was missing for one article (Hartwig and Bond Jr., 2011). In the second version, we included the columns "Entry in reference", "Analysis description", "*I*²", and "Test for publication bias" and removed a column in which the type of effect size was recoded. Moreover, we discovered coding errors for the column "Publication bias?". Specifically, several articles which did not test for publication bias were coded as having no publication bias. This has been changed to NA for the articles: Else-quest et al. (2012); Hagger et al. (2010); van Zomeren et al. (2008); Postmes and Spears (1998); Ambady and Rosenthal (1992); and Raz and Raz (1990). In addition, we changed the recoded publication bias variable to remain NA if information regarding publication bias was not reported in the article, instead of coding this as 2.

Object name

The name of the file is "Data 1990–2013 with tau values.xlsx".

Reading and cleaning the data

The following R code can be used to read in the data and clean it:

```
require(xlsx) #to read in xlsx files
setwd() #specify directory where the data is
located
dat <- read.xlsx("Data 1990-2013 with tau
values.xlsx", sheetIndex = 1)
#remove missing estimates:
dat.noNA <- dat[-c(which(is.na(dat$tau.2))), ]
attach(dat.noNA)
sum(tau.2 == 0) #estimates equal to 0
```

Some heterogeneity estimates are missing for articles in which meta-analyses were conducted for multiple variables. In these cases, it sometimes happened that only one study was found for some reported variables and thus only one effect size estimate is available. These effect sizes have been included in the data set, but naturally there is no corresponding heterogeneity estimate available. An example is the article by DeNeve and Cooper (1998). In addition, heterogeneity estimates were sometimes missing despite the fact that multiple effect sizes were available. This occurred once in the article by DeNeve and Cooper (1998), who did not provide an explanation. It occurred multiple times in the article by Mathieu and Zajac (1990) when the between study heterogeneity was completely attributable to statistical artifacts.

A total of 90 heterogeneity estimates are equal to zero. When there exists no between-study variance, a fixed-effect meta-analysis is appropriate. However, some articles reported random-effects meta-analyses by default or reported both fixed-effect and random-effects meta-analyses, in which case the heterogeneity estimate can be zero.

Data collectors

Sara van Erp (Tilburg University) collected the data.

Language

English.

License

CC0 1.0 Universal.

Embargo

The data set is not under embargo.

Repository location

The data set is published in the Open Science Framework repository, available at <https://osf.io/wyhve/>.

Publication date

The data set was originally published on 29/11/2015. The second version was published on 08/04/2017 and the final version was published on 17/07/2017.

Reuse potential

The data set provides an overview of between-study heterogeneity estimates based on 61 meta-analyses in psychology. The main application of this data set is to facilitate the construction of an informed prior for the between-study

heterogeneity in Bayesian meta-analysis, which has been done in [10] for heterogeneity in mean difference effect sizes. Researchers interested in a Bayesian meta-analysis on correlations can use the heterogeneity estimates for the correlation effect sizes to construct an informed prior. More generally, the description of between-study heterogeneity available in this data set is useful as reference. For example, researchers can compare an obtained τ^2 estimate to the distribution of estimates and assess the relative size of the between-study heterogeneity in their meta-analysis.

In addition to the τ^2 estimates of the between-study heterogeneity, the data set contains information on the number of studies in each meta-analysis, the Q -statistic and presence of publication bias. This information can be used for example in Monte Carlo simulation studies to determine a realistic range of population values for these variables. Moreover, the data set can be used by researchers interested in the effects of publication bias on heterogeneity estimates or the Q -test. Finally, as new methods to test for publication bias become available, researchers might want to use the information on publication bias as reference point to compare the new methods to.

Additional Files

The additional files for this article can be found as follows:

- **Appendix.** List of included meta-analyses (all studies are published in *Psychological Bulletin*). DOI: <https://doi.org/10.5334/jopd.33.s1>
- **Data.** Data set containing 705 between-study heterogeneity estimates from meta-analyses published in *Psychological Bulletin* from 1990-2013. DOI: <https://doi.org/10.5334/jopd.33.s2>

Competing Interests

The authors have no competing interests to declare.

References

1. Pullenayegum, E M 2011 "An informed reference prior for between-study heterogeneity in meta-analyses of binary outcomes". *Statistics in Medicine*, 30(26): 3082–3094. DOI: <https://doi.org/10.1002/sim.4326>
2. Rhodes, K M, Turner, R M and Higgins, J P 2015 "Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data". *Journal of Clinical Epidemiology*, 68(1): 52–60. DOI: <https://doi.org/10.1016/j.jclinepi.2014.08.012>
3. Turner, R M, Davey, J, Clarke, M J, Thompson, S G and Higgins, J P 2012 "Predicting the extent of heterogeneity in meta-analysis, using empirical data from the cochrane database of systematic reviews". *International Journal of Epidemiology*, 41(3): 818–827. DOI: <https://doi.org/10.1093/ije/dys041>
4. Lipsey, M W and Wilson, D B 1993 "The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis". *American Psychologist*, 48(12): 1181–1209. DOI: <https://doi.org/10.1037/0003-066X.48.12.1181>

5. **Schmidt, F L, Oh, I-S and Hayes, T L** 2009 “Fixed-versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results”. *British Journal of Mathematical and Statistical Psychology*, 62(1): 97–128. DOI: <https://doi.org/10.1348/000711007X255327>
6. “Psychological bulletin.” <http://www.apa.org/pubs/journals/bul/?tab=1> Accessed: April 6, 2017.
7. **Viechtbauer, W** 2005 “Bias and efficiency of meta-analytic variance estimators in the random-effects model”. *Journal of Educational and Behavioral Statistics*, 30(3): 261–293. DOI: <https://doi.org/10.3102/10769986030003261>
8. **DerSimonian, R and Laird, N** 1986 “Meta-analysis in clinical trials”. *Controlled Clinical Trials*, 7(3): 177–188. DOI: [https://doi.org/10.1016/0197-2456\(86\)90046-2](https://doi.org/10.1016/0197-2456(86)90046-2)
9. **Higgins, J P T and Thompson, S G** 2002 “Quantifying heterogeneity in a meta-analysis”. *Statistics in Medicine*, 21(11): 1539–1558. DOI: <https://doi.org/10.1002/sim.1186>
10. **Gronau, Q F, van Erp, S, Heck, D W, Cesario, J, Jonas, K J and Wagenmakers, E-J** 2017 “A bayesian model-averaged meta-analysis of the power pose effect with informed and default priors: the case of felt power”. *Comprehensive Results in Social Psychology*, 2(1): 123–138. DOI: <https://doi.org/10.1080/23743603.2017.1326760>

How to cite this article: van Erp, S, Verhagen, J, Grasman, R P P P and 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: 4, DOI: <https://doi.org/10.5334/jopd.33>

Published: 17 August 2017

Copyright: © 2017 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.