Meta-analyzing an original study and replication

Robbie C.M. van Aert

December 13, 2019

European Research Council
Established by the European Commission

Tilburg University

Chapters of my dissertation

- Chapters 2, 3, and 5 $\rightarrow$ correcting for publication bias in a meta-analysis (*p*-uniform, *p*-uniform*)

- Chapter 4 $\rightarrow$ Meta-meta-analysis on publication bias in psychology and medicine

- Chapter 6 and 7 $\rightarrow$ meta-analyzing original study and replication (Hybrid, snapshot)

- Chapter 8 $\rightarrow$ Multi-step estimator for estimating between-study variance in a meta-analysis (together with Dr. Dan Jackson)

- Chapter 9 $\rightarrow$ Assessing properties of methods for constructing a confidence interval for the between-study variance (together with Dr. Wolfgang Viechtbauer)

Overview

1. Bias in published literature
2. Replications and meta-analysis: The problem
3. Hybrid meta-analysis method
4. Snapshot Bayesian Hybrid Meta-Analysis method
5. Application: Replicability projects
6. Software
7. Conclusion and discussion

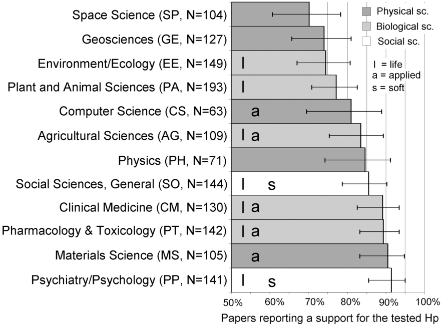
1. Bias in published literature

- Overwhelming evidence for bias in published literature

- ≈90% of main hypotheses are significant in psychology

- But this is not in line with average statistical power (about 20-50%)

- Consequences:
  - Overestimation
  - False impression



Adapted from Fanelli (2010)

2. Replications and meta-analysis: The problem

- Example of a common problem (independent samples t-test):

	Cohen's d	t-statistic
Original	0.5	$t(78) = 2.24$, $p = .028$
Replication	0.23	$t(170) = 1.5$, $p = .135$

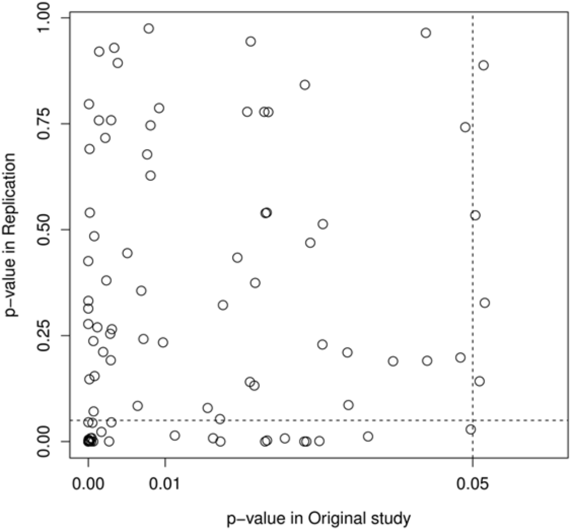
What to conclude?!

Questions considered relevant:

- Does an effect exist? (0 or not)
- What is the magnitude of effect size? (best guess)

2. Replications and meta-analysis: The problem

- Distribution of *p*-values in Reproducibility Project: Psychology
  - Significant original and nonsignificant replication in 63.9%



2. Replications and meta-analysis: The problem

- Significant results are overrepresented in the literature

- Published effect sizes are therefore most probably overestimated

- Replicability projects in psychology (RPP) and economics (EE-RP) confirmed that effect sizes are overestimated:
  - RPP: $r = 0.403$ vs. 0.197
  - EE-RP: $r = 0.506$ vs. 0.303

2. Replications and meta-analysis: The problem

- Significant results are overrepresented in the literature

- Published effect sizes are therefore most probably overestimated

- Replicability projects in psychology (RPP) and economics (EE-RP) confirmed that effect sizes are overestimated:
  - RPP: $r = 0.403$ vs. 0.197
  - EE-RP: $r = 0.506$ vs. 0.303

- **Conclusion:** We should take statistical significance of original study into account

3. Hybrid method: Basic idea

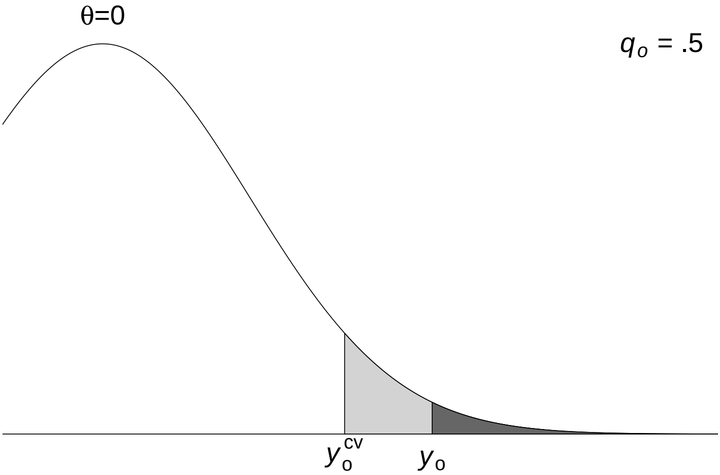
- Distribution of *p*-values at the true effect size is uniform

- Original study is significant, so we compute a conditional probability:

$$q_o = \frac{P(y \geq y_o; \theta)}{P(y \geq y_o^{cv}; \theta)}$$

3. Hybrid method: Basic idea

- Illustration computing conditional probability



3. Hybrid method: Basic idea

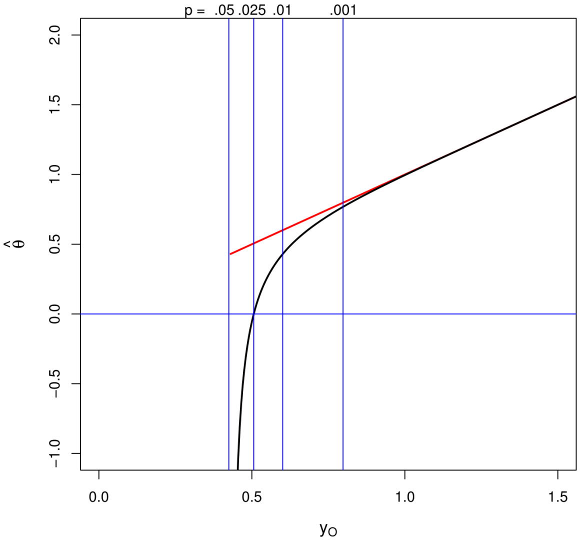
- Distribution of p-values at the true effect size is uniform

- Original study is significant, so we compute a conditional probability:

$$q_o = \frac{P(y \geq y_o; \theta)}{P(y \geq y_o^{cv}; \theta)}$$

- Effect size estimate is obtained when $q_o = 0.5$

3. Hybrid method: Basic idea



3. Hybrid method: Basic idea

- Distribution of p-values at the true effect size is uniform

- Original study is significant, so we compute a conditional probability:

$$q_o = \frac{P(y \geq y_o; \theta)}{P(y \geq y_o^{cv}; \theta)}$$

- "Normal" probability for replication because replication does not have to be significant:

$$q_r = P(y \geq y_r; \theta)$$

3. Hybrid method: Basic idea

- Combined effect size estimate is obtained when sum of conditional probabilities equals 1

- We can also create a confidence interval and test null-hypothesis of no effect

- Assumptions:
  - Original study is statistically significant
  - Both studies estimate the same effect (fixed-effect)
  - No questionable research practices

3. Variants of Hybrid method

- Estimates of hybrid method can become highly negative if (conditional) probabilities are close to 1

Variants of Hybrid method:

- Set effect size estimate to 0 if mean of probabilities under null-hypothesis is larger than 0.5 $\rightarrow$ Hybrid^0 method

- Use effect size estimate of replication if *p*-value in original study is larger than $\alpha/2$ $\rightarrow$ Hybrid^R method

3. Method of computations

- 1,000 equally spaced probabilities given significant original study and 1,000 equally spaced probabilities for replication:

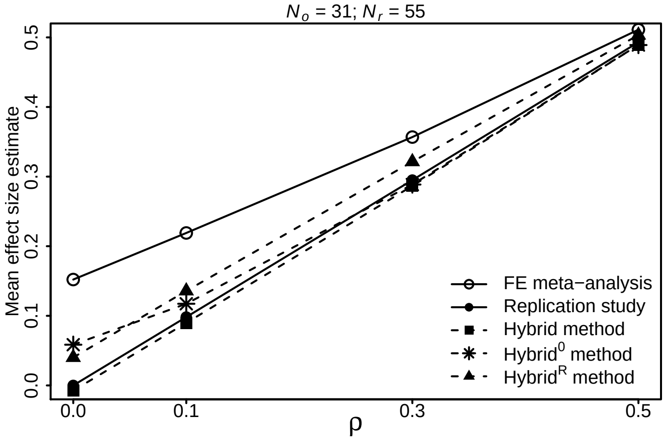
$$1{,}000 \times 1{,}000 = 1{,}000{,}000 \text{ combinations}$$

- Independent variables:
  - True effect size $\rho = 0$; 0.1; 0.3; 0.5
  - Sample sizes in original study and replication: 31; 55; 96

- Dependent variables:
  - Mean and median effect size estimate
  - Root mean square error
  - Coverage probability
  - Type-I error rate and statistical power

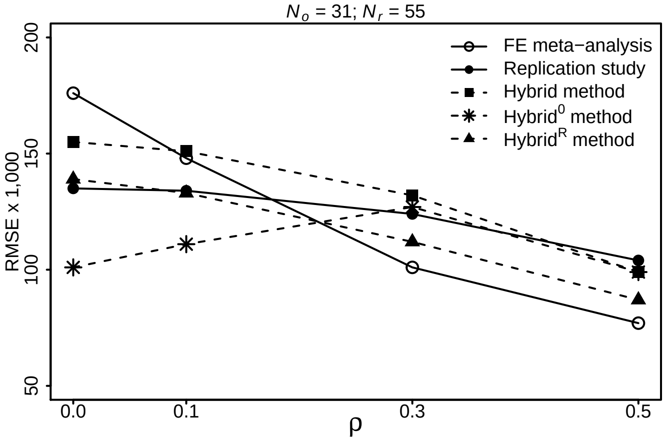
- Methods: Hybrid methods, fixed-effect meta-analysis, and only considering replication

3. Results: Mean ES estimate



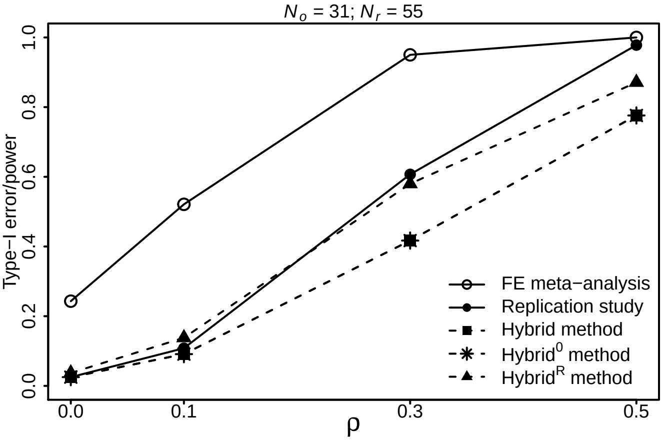
- FE meta-analysis substantially overestimates ρ
- Replication and Hybrid method slightly underestimate ρ

3. Results: RMSE



- Precision of FE meta-analysis is higher than other methods
- Hybrid^0 has the lowest RMSE if $\rho = 0$ or 0.1

3. Results: Type-I error rate and power



- Type-I error rate of FE meta-analysis too high
- Power is deceivingly high for FE meta-analysis
- Replication has more power than Hybrid methods in some conditions

3. Guidelines based on results

1. *When uncertain about true effect size...*
   a. and $N_r > N_o \rightarrow$ use only replication data
   b. and $N_r \leq N_o \rightarrow$ use Hybrid^R

2. *When suspecting zero or small true effect size* $\rightarrow$ Hybrid^R

3. *When suspecting medium or large true effect size* $\rightarrow$ fixed-effect meta-analysis

4. Snapshot method

- **Snapshot** Bayesian Hybrid Meta-Analysis Method
  - Assume four effect sizes (zero, small, medium, large) → *snapshots*
- Snapshot **Bayesian** Hybrid Meta-Analysis Method
  - Compute posterior probability of these four effects → *Bayesian*
- Snapshot Bayesian **Hybrid** Meta-Analysis Method
  - Take statistical significance of original study into account → *hybrid*
- Snapshot Bayesian Hybrid **Meta-Analysis** Method
  - Combine original study with replication → *meta-analysis*

4. Snapshot method: Basic idea

- Density of the replication is "normal" pdf because no selection:

$$f_r = f(y = y_r; \theta)$$

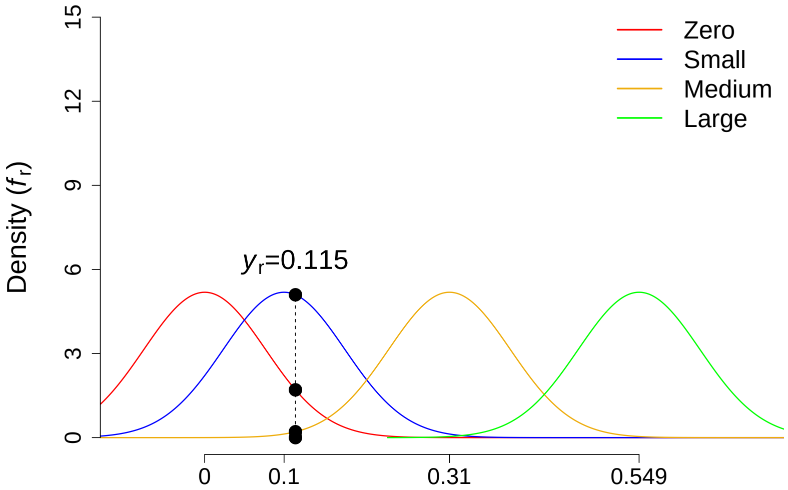
- Density of the original study is pdf *conditional on effect size being statistically significant*:

$$f_o = \frac{f(y = y_o; \theta)}{P(y \geq y_o^{cv}; \theta)}$$

- Assumptions:
  - Original study is statistically significant
  - Both studies estimate the same effect (fixed-effect)
  - No questionable research practices

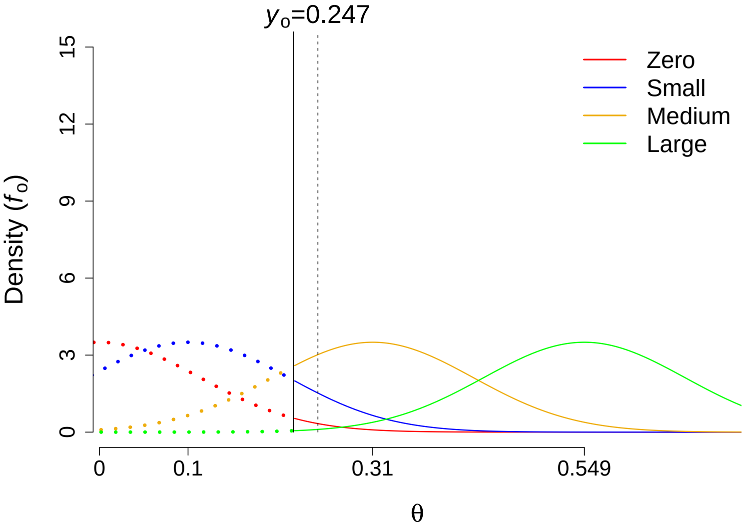
4. Snapshot method: Basic idea

- Densities replication: $d = 0.23$, $t(170) = 1.5$, $p = 0.135$



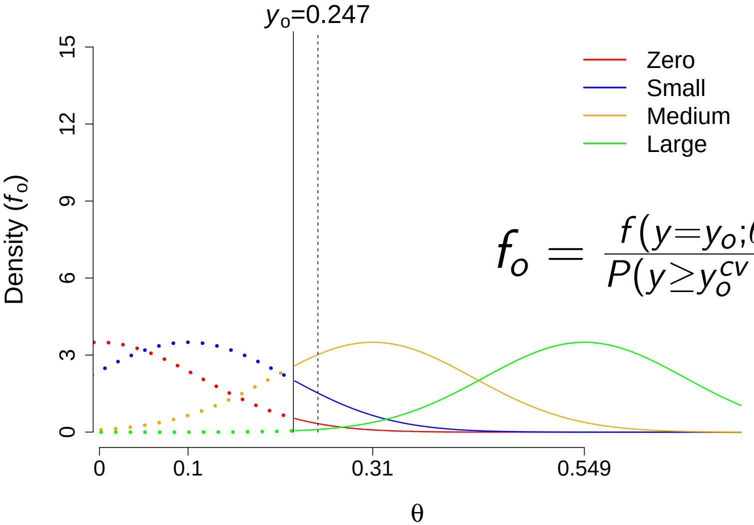
4. Snapshot method: Basic idea

- Densities original study (naïve): $d = 0.5$, $t(78) = 2.24$, $p = 0.028$



4. Snapshot method: Basic idea

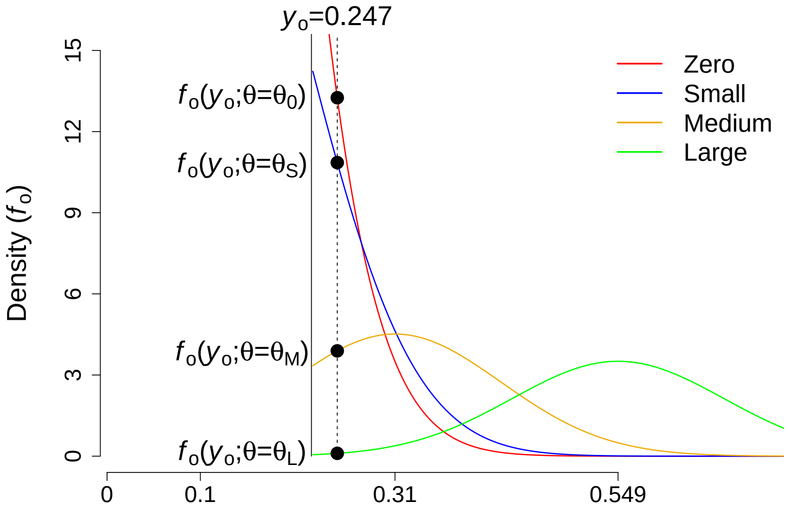
- Densities original study (naïve): $d = 0.5$, $t(78) = 2.24$, $p = 0.028$



$$f_o = \frac{f(y=y_o;\theta)}{P(y \geq y_o^{cv};\theta)}$$

4. Snapshot method: Basic idea

- Densities original study: $d = 0.5$, $t(78) = 2.24$, $p = 0.028$



4. Snapshot method: Basic idea

- Combined likelihood:

$$L(\theta) = f_o(\theta) \times f_r(\theta)$$

- Posterior probabilities assuming a uniform prior for each snapshot are computed with:

$$\pi_x = \frac{L(\theta = x)}{L(\theta = \theta_0) + L(\theta = \theta_S) + L(\theta = \theta_M) + L(\theta = \theta_L)}$$

4. Snapshot method: Basic idea

- Combined likelihood:

$$L(\theta) = f_o(\theta) \times f_r(\theta)$$

- Posterior probabilities assuming a uniform prior for each snapshot are computed with:

$$\pi_x = \frac{L(\theta = x)}{L(\theta = \theta_0) + L(\theta = \theta_S) + L(\theta = \theta_M) + L(\theta = \theta_L)}$$

- Advantages of the method:
  - Easy and insightful
  - Easy (re)computation posterior for other (than uniform) prior:

$$\pi_x^* = \frac{p_x \pi_x}{p_0 \pi_0 + p_S \pi_S + p_M \pi_M + p_L \pi_L}$$

5. Application

- Initiatives to study the replicability of psychological research

- **RPP:** Studies from JPSP, Psychological Science, and Journal of Experimental Psychology: 67 out of 100 studies were included

- "High-powered" replication of a key effect

- Effect sizes were transformed to correlation coefficients

5. Application: Hybrid method

Effect size estimation:

		Overall	JEP: LMC	JPSP	PSCI, cog.	PSCI, soc.
# study-pairs		67	20	18	13	16
Mean (SD)	FE	0.322 (.229)	0.416 (.205)	0.133 (.083)	0.464 (.221)	0.300 (.241)
	Replication	0.199 (.280)	0.291 (.264)	0.026 (.097)	0.289 (.365)	0.206 (.292)
	Hybrid	0.250 (.263)	0.327 (.287)	0.071 (.087)	0.388 (.260)	0.245 (.275)
	$Hybrid^0$	0.266 (.242)	0.353 (.237)	0.080 (.075)	0.400 (.236)	0.257 (.259)
	$Hybrid^R$	0.268 (.254)	0.368 (.241)	0.083 (.093)	0.394 (.272)	0.247 (.271)

5. Application: Hybrid method

Test of null-hypothesis of no effect:

		Overall	JEP: LMC	JPSP	PSCI, cog.	PSCI, soc.
%Significant results	FE	70.1%	90%	44.4%	92.3%	56.2%
	Replication	34.3%	50%	11.1%	46.2%	31.2%
	Hybrid	28.4%	45%	11.1%	30.8%	25%
	Hybrid^0	28.4%	45%	11.1%	30.8%	25%
	Hybrid^R	34.3%	55.5%	16.7%	38.5%	25%

- **Conclusions:**
  - Effect size estimates of hybrid methods smaller than FE meta-analysis but larger than replication
  - Replication and hybrid methods yielded less significant results than FE meta-analysis

5. Application: Snapshot method

- Probability of strong evidence ($\pi_x >.75$; BF >3) using snapshot method

	Zero	Small	Medium	Large	Unknown
EE-RP	0	0.062	0.312	0.438	0.188
RPP	0.134	0.03	0.045	0.164	0.627

- **Conclusions:**
  - Studied effects larger in EE-RP than in RPP
  - Only few studies have strong evidence for zero effect in RPP (13.4%)
  - Often not enough information for determining magnitude of effect size in RPP (62.7%)

6. Software

- Hybrid and snapshot method are both implemented in the R package `puniform`

- Web applications are available for researchers unfamiliar with R:
  - https://rvanaert.shinyapps.io/snapshot/
  - https://rvanaert.shinyapps.io/hybrid/

7. Conclusion and discussion

- Methods *should* take statistical significance of original study into account

- Analyzing replicability projects shows that adjusting for statistically significant original study influences the conclusions

- Determining sample size of replication with snapshot method akin to computing required sample size with power analysis

7. Conclusion and discussion

- Methods *should* take statistical significance of original study into account
- Analyzing replicability projects shows that adjusting for statistically significant original study influences the conclusions
- Determining sample size of replication with snapshot method akin to computing required sample size with power analysis
- Future research:
  - Extend methods such that these can deal with multiple original studies and replications
  - Implement intervals of effects sizes rather than discrete values as snapshots
  - Effect size estimation with the Bayesian method

Thank you for your attention

www.robbievanaert.com

www.metaresearch.nl

van Aert, R. C. M., & van Assen, M. A. L. M. (2018). Examining reproducibility in psychology: A hybrid method for combining a statistically significant original study and a replication. Behavior Research Methods, 50(4), 1515-1539. doi:10.3758/s13428-017-0967-6

van Aert, R. C. M., & van Assen, M. A. L. M. (2017). Bayesian evaluation of effect size after replicating an original study. PLOS ONE, 12(4), e0175302. doi:10.1371/journal.pone.0175302