Analyzing Data of a Multi-Lab Replication Project with Individual Participant Data Meta-Analysis: A Tutorial

Robbie C.M. van Aert

January 7, 2021

European Research Council

Established by the European Commission

Overview

1. Multi-lab replication projects
2. Individual Participant Data (IPD) meta-analysis
3. Two-step vs. one-step IPD meta-analysis
4. Example: RRR of McCarthy et al.
5. Discussion

Multi-lab replication projects

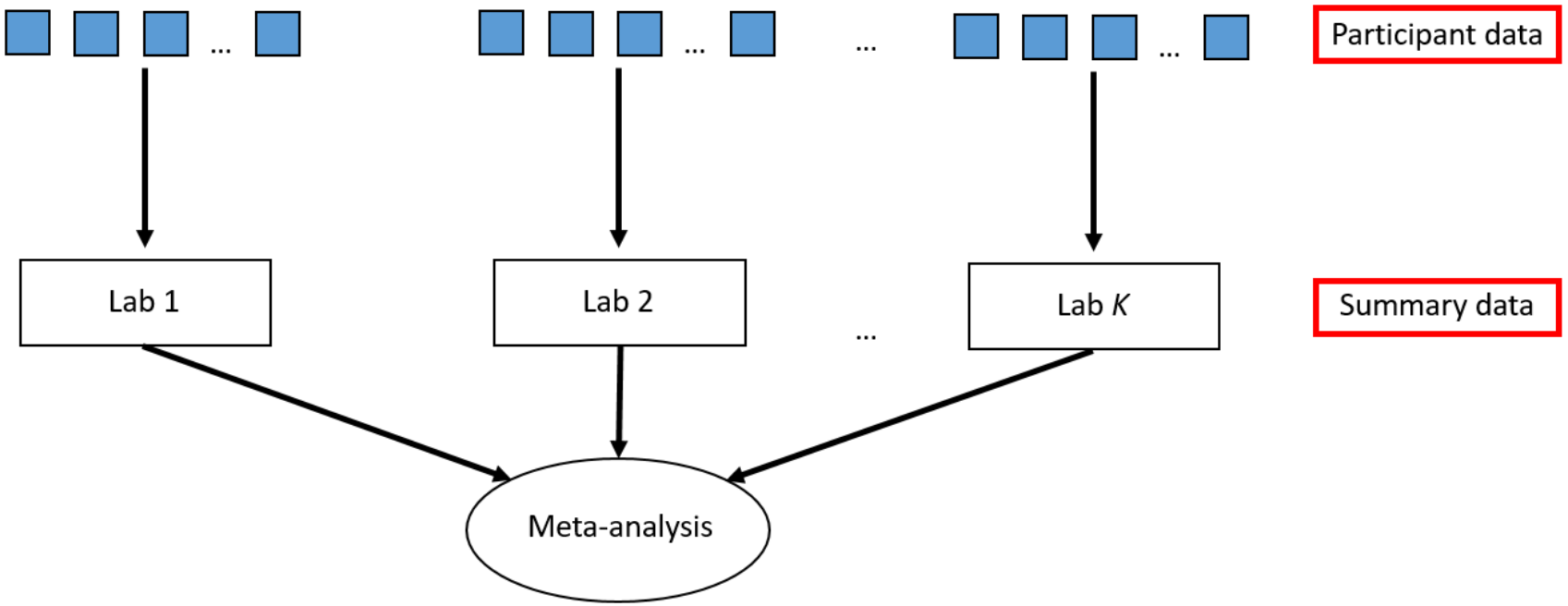
- Prominent effects are replicated in multiple labs to study
  1. Replicability $\rightarrow$ can the effect be replicated?
  2. Robustness $\rightarrow$ does the effect depend on contextual factors?

- Examples are Registered Replication Reports (RRRs) and Many Labs projects

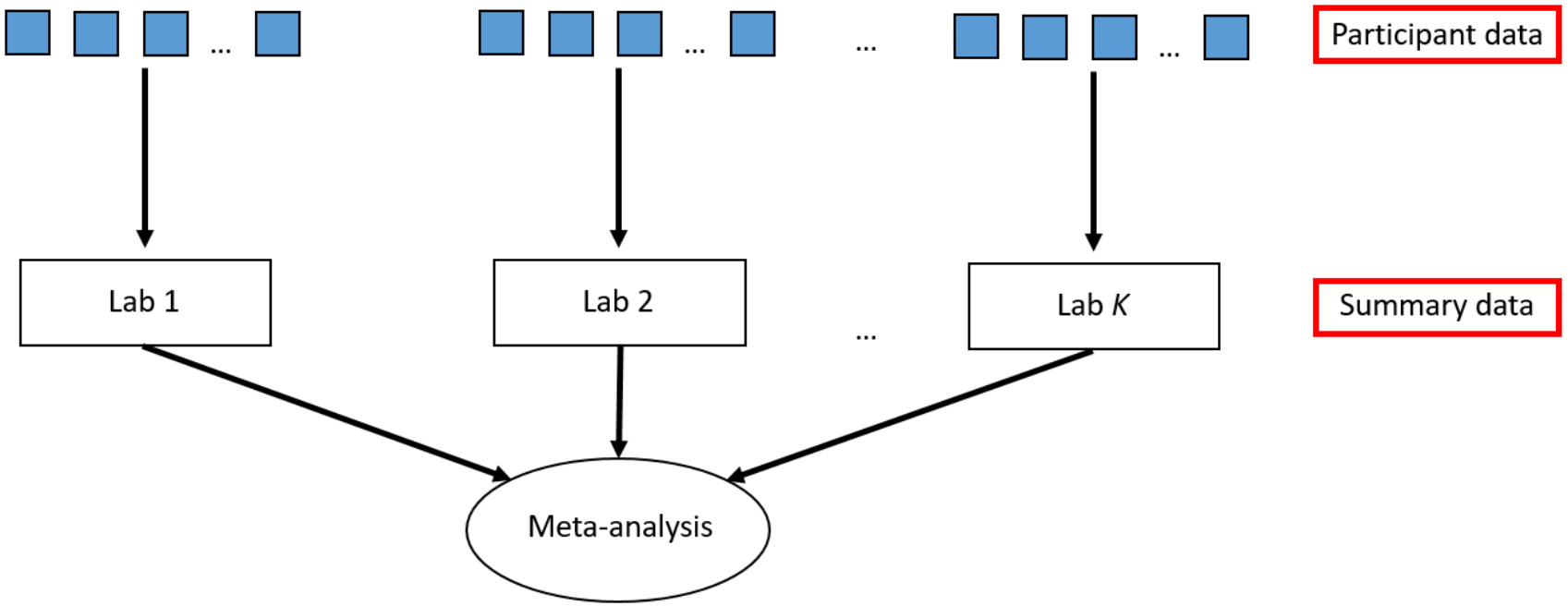
- Twelve RRRs are currently published in *Perspectives on Psychological Science* and *AMPPS*

- Sixty effects were replicated in Many Labs 1, 2, 3, and 5

Data multi-lab replication project



Data multi-lab replication project



- 70 out of 72 (97.2%) published multi-lab projects analyzed summary data in their primary analysis

Individual Participant Data (IPD) meta-analysis

- IPD meta-analysis models are multilevel models applied to participants who are nested in studies
- Advantages of IPD meta-analysis over conventional meta-analysis:
  - Statistical power is generally larger

Individual Participant Data (IPD) meta-analysis

- IPD meta-analysis models are multilevel models applied to participants who are nested in studies

- Advantages of IPD meta-analysis over conventional meta-analysis:
  - Statistical power is generally larger
  - Studies that use the same data can be identified

Individual Participant Data (IPD) meta-analysis

- IPD meta-analysis models are multilevel models applied to participants who are nested in studies

- Advantages of IPD meta-analysis over conventional meta-analysis:
  - Statistical power is generally larger
  - Studies that use the same data can be identified
  - Data analysis in the studies can be standardized (e.g., handling missing data, outlier removal)

Individual Participant Data (IPD) meta-analysis

- IPD meta-analysis models are multilevel models applied to participants who are nested in studies

- Advantages of IPD meta-analysis over conventional meta-analysis:
  - Statistical power is generally larger
  - Studies that use the same data can be identified
  - Data analysis in the studies can be standardized (e.g., handling missing data, outlier removal)
  - Transforming effect sizes is not needed

Individual Participant Data (IPD) meta-analysis

- IPD meta-analysis models are multilevel models applied to participants who are nested in studies

- Advantages of IPD meta-analysis over conventional meta-analysis:
  - Statistical power is generally larger
  - Studies that use the same data can be identified
  - Data analysis in the studies can be standardized (e.g., handling missing data, outlier removal)
  - Transforming effect sizes is not needed
  - Participant level moderators can be included to explain heterogeneity in effect size

Two-step vs. one-step IPD meta-analysis

- Both approaches allow drawing conclusions at the participant level

- **Two-step:** effect sizes are computed per lab and synthesized using conventional meta-analysis models
  - Advantage: Similar to conventional meta-analysis
  - Disadvantage: Low statistical power

- **One-step:** participant data are modeled directly using a multilevel approach
  - Advantages: More flexible model and larger statistical power
  - Disadvantage: More complex $\rightarrow$ convergence problems

Example: RRR on assimilative priming

- McCarthy et al. (2018) replicated the study by Scrull and Wyer (1979) on assimilative priming

- Assimilative priming refers to the idea that "exposure to priming stimuli causes subsequent judgments to incorporate more of the qualities of the primed construct"

- Procedure replicated experiment:
  - Participants performed a sentence construction task with 20% or 80% of the sentences describing hostile behavior
  - Participants read a vignette about a person who behaved in an ambiguously hostile way and rated whether he was perceived as hostile

Example: RRR on assimilative priming

- **Hypothesis:** Participants exposed to a larger number of sentences describing hostile behavior would rate the person's behavior as more hostile
- Raw mean difference was the effect size measure of interest
- A positive difference indicates that the hostility rating was larger in the 80% condition
- 22 labs participated yielding a total sample size of 7,373

Example: Two-step IPD

- **First step:** A linear regression model is fitted to the data of each lab,

$$y_j = \phi + \theta x_j + \epsilon_j$$

ϕ = fixed lab effect
θ = treatment effect
x_j = dummy variable (0 = 20%, 1 = 80% condition)

- In R: `lm(y ~ x)`

Example: Two-step IPD

- **First step:** A linear regression model is fitted to the data of each lab,

$$y_j = \phi + \theta x_j + \epsilon_j$$

ϕ = fixed lab effect
θ = treatment effect
x_j = dummy variable (0 = 20%, 1 = 80% condition)

- In R: `lm(y ~ x)`

- **Second step:**
  - $\hat{\theta}$ obtained in the first step for each lab are meta-analyzed using a conventional meta-analysis model
  - In R using `metafor` package:

```
rma(yi = theta_hat, vi = vi_theta_hat)
```

Example: One-step IPD

- A *single* multilevel model is fitted to the data,

$$y_{ij} = \phi_i + \theta_i x_{ij} + \epsilon_{ij}$$

- A controversial decision is whether the lab effect (ϕ_i) are treated as fixed or random parameters
  - Fixed: the lab effect is estimated for each lab $\rightarrow$ large number of parameters
  - Random: lab effects are assumed to be sampled from a normal distribution

Example: One-step IPD

- A *single* multilevel model is fitted to the data,

$$y_{ij} = \phi_i + \theta_i x_{ij} + \epsilon_{ij}$$

- A controversial decision is whether the lab effect (ϕ_i) are treated as fixed or random parameters
  - Fixed: the lab effect is estimated for each lab $\rightarrow$ large number of parameters
  - Random: lab effects are assumed to be sampled from a normal distribution

- Results with random lab effects will be shown as these allow generalizing the results to the population of effects

- In R using the `lme4` package: `lmer(y ~ x + (x | lab))`

Example: Results

	$\hat{\mu}$ (SE)	(95% CI)	$\hat{\tau}^2$	(95% CI)
RE MA	0.083 (0.04)	(0.004;0.161)	0.006	(0;0.043)
Two-step	0.082 (0.04)	(0.004;0.161)	0.006	(0;0.043)
One-step	0.09 (0.038)	(0.017;0.164)	0.002	-

- Results of random-effects meta-analysis match those of McCarthy et al. :-)
- Hardly any difference between estimates and CIs of different approaches, but CI of one-step is the smallest

Example: Two-step IPD with moderator

- **First step:** A linear regression model containing the moderator is fitted to the data of each lab,

$$y_j = \phi + \alpha w_j + \theta x_j + \gamma w_j x_j + \epsilon_j$$

$\alpha =$ main effect of the moderator w
$\gamma =$ interaction between treatment and moderator

- In R: `lm(y ~ x + age + x:age)`

- **Second step:**
  - $\hat{\gamma}$ obtained in the first step for each lab are meta-analyzed using a conventional meta-analysis model
  - In R using `metafor` package:

```
rma(yi = gamma_hat, vi = vi_gamma_hat)
```

Example: One-step IPD with moderator

- One-step IPD can disentangle the within *and* between lab interaction between the treatment and moderator

- We need group-mean centering for this $\rightarrow$ subtracting the lab's mean from the moderator variable,

$$y_{ij} = \phi_i + \alpha_i w_{ij} + \theta_i x_{ij} + \gamma_W x_{ij}(w_{ij} - m_i) + \gamma_B x_{ij} m_i + \epsilon_{ij}$$

m_i = mean score on moderator variable of the ith lab
γ_W = within-lab interaction between treatment and moderator
γ_B = between-lab interaction between treatment and moderator

Example: One-step IPD with moderator

- One-step IPD can disentangle the within *and* between lab interaction between the treatment and moderator

- We need group-mean centering for this $\rightarrow$ subtracting the lab's mean from the moderator variable,

$$y_{ij} = \phi_i + \alpha_i w_{ij} + \theta_i x_{ij} + \gamma_W x_{ij}(w_{ij} - m_i) + \gamma_B x_{ij} m_i + \epsilon_{ij}$$

m_i = mean score on moderator variable of the ith lab
γ_W = within-lab interaction between treatment and moderator
γ_B = between-lab interaction between treatment and moderator

- In R using the `lme4` package:

```
lmer(y ~ x + (x | lab) + age + I(age-age_gm):x + age_gm:x)
```

Example: Results moderator analysis

		Estimate (SE)	(95% CI)	$\hat{\tau}^2$	(95% CI)
RE MR	Intercept	-0.921 (0.812)	(-2.512;0.671)	0.005	(0;0.043)
	Mean age	0.05 (0.04)	(-0.029;0.128)	0.005	(0;0.043)
Two-step	Age	0.053 (0.024)	(0.007;0.1)	0	(0;0.011)
One-step	Intercept	8.264 (0.353)	(7.57;8.951)		
	x	-0.791 (0.814)	(-2.318;0.82)	0.003	-
	Age	-0.064 (0.017)	(-0.096;-0.03)		
	Age within	0.05 (0.024)	(0.003;0.096)	0.003	-
	Age between	0.044 (0.04)	(-0.036;0.119)	0.003	-

- No effect of mean age in meta-regression model
- Interaction between the treatment and age *within* but not *between* labs according to two-step and one-step IPD

Discussion

- Applying conventional meta-analysis to data of multi-lab replication projects is suboptimal
- Especially one-step IPD meta-analysis is ideal for analyzing these data
- However, convergence issues may arise in one-step IPD meta-analysis $\rightarrow$ simplify model or use two-step IPD meta-analysis

Discussion: Extensions

- IPD meta-analysis can also be used in internal meta-analyses
- Model flexibility of one-step meta-analysis $\rightarrow$ extra random-effects
- Hopefully, sharing participant data becomes the norm and IPD meta-analysis can regularly be applied

Thank you for your attention

www.robbievanaert.com

www.metaresearch.nl

Preprint paper:

https://osf.io/preprints/metaarxiv/9tmua/