Project proposal: Explanations for unexplained heterogeneity in meta-analysis

Anton Olsson Collentine

2021-05-17

* When studies fail to replicate heterogeneity is often raised as a potential explanation.
* This explanation is supported by most meta-analyses in psychology reporting large heterogeneity in the studied effect (Stanley, van Erp). Even in medicine, where heterogeneity estimates tend to be smaller (Ioannidis), the majority of meta-analyses appear to find heterogeneity (between 50-75%; Viechtbauer referencing Field; Higgins; Lipsey).
* Some authors have argued that heterogeneity is inevitable in meta-analysis (Higgins), an argument which finds support in recent research demonstrating that substantial heterogeneity between studies can arise even from apparently minor differences in data processing (multiverse papers),
* Moreover, most preregistered multi-lab projects in psychology, where data collection and processing is identical across studies, seem to find little or no heterogeneity (Olsson-Collentine et al.)
* It seems like in most situations (the exception being multi-lab projects or meta-analyses with very strict inclusion criteria) we should expect meta-analyses to have heterogeneity
* However, heterogeneity estimates in meta-analysis are almost invariably discussed as representing differences in true effect sizes between included studies (e.g., refs)
* That is, arising from moderator effects
* However, heterogeneity estimates in meta-analysis may also arise from statistical artefacts.
* These may derive from, for example, the estimation procedure (von Hippel), differential measurement error (Olsson-Collentine et al), publication bias (Augusteijn), and even arbitrary decisions in the research process (multiverse papers)
* Fundamentally, anything that affects effect size estimates may also affect statistical heterogeneity estimates.
* As such, researchers should make a concerted effort to rule out artefactual explanations of statistical heterogeneity before examining moderators
* In this paper, we describe several explanations for unexplained heterogeneity in meta-analysis, examine their general effect on heterogeneity estimates and how to rule them out.
* We first provide a short definition of heterogeneity and how it is quantified.

#### What is heterogeneity

* It may be useful to distinguish between clinical heterogeneity (differences in contextual factors between studies) and statistical heterogeneity (a statistical estimate of the sensitivity to differences in contextual factors between studies).
* A statistical definition of heterogeneity may make it more clear why ‘heterogeneity’ cannot be assumed to directly represent the differences in true effects between included studies.
* We can statistically define heterogeneity as: the residual variance between effect sizes after taking into account sampling error.
* Mention something about estimation, but probably don’t go into details (provide references).
* Mention different estimators (I2, Q, Tau2), but discuss them further when talking about explanations

## Methods

#### Potential empirical aspects to add

1. Strictness of inclusions medicine vs psychology
   * How many papers do meta-analysts exclude from their initial search?
     + Check PRISMA flowcharts in meta-analyses.
   * And/or compare the number of inclusion criteria
2. Literature review of different QRPs that affect heterogeneity (“poor methodological quality”/ “small study effects”)
3. Moderators (% explained variance vs. power)
   * Perhaps we can just do a smallish review, e.g., look at Robbie’s meta-meta data?
4. Check how often I2 used over tau (might already be some paper who has done this)
5. Check how often sample sizes in primary studies are large enought to lead to substantial I2 heterogeneity
6. Check how often reliability is available in primary studies (or refer to e.g., Flake et al)
7. Use the meta-multiverse data to demonstrate the effect of arbitrary decisions on heterogeneity

## Explanations for heterogeneity

* Here we first list potential artefactual explanations and how they might be examined. We then shortly discuss examining potential moderators of true heterogeneity, but primarily refer the reader to the several excellent papers available on this topic.

#### Explanation 1: estimation artefacts

1. I2 -> heterogeneity increases with primary study sample sizes
   * Rule out by using tau instead
2. Central chi-square distribution (Q->I2/tau) can lead to heterogeneity (von Hippel)
   * If heterogeneity is large enough, not a sufficient explanation = can rule out

#### Explanation 2: measurement artefacts

* Differential measurement error
  + Can examine how large the effect would be on estimates if we know reliability in all primary studies? = rule out as sufficient explanation if heterogeneity too large
* Range restrictions? (Schmidt & Hunter, 2015; cited in Olsson-Collentine et al.)
* “poor methodological quality”

#### Explanation 3: arbitrary data processing and model decisions

* Multiverse papers
  + Can never rule out as explanation (maybe with access to raw data?)

#### Explanation 4: QRPs

* Publication bias (Hilde)
  + Can try to rule out with inclusion of grey literature, publication bias tests
* Selective reporting (including p-hacking) ((“small study bias”)
  + Different forms of p-hacking?
  + Can rule out only by checking preregistration

#### Explanation 5: systematic variance (moderators)

* Short explanation

## Summary & Conclusion

* There are many potential explanations for why the purportedly same true effect size differs between studies and results in heterogeneity in meta-analysis.
* Before examining moderators as an explanation, researchers should make a concerted effort to rule out artefactual explanations such as those discussed in this paper.