**Combining data**

By: Wilhelmina van Dijk & Christopher Schatschneider

Florida State University

For some research questions, it is better to have large sample sizes. Instead of going out to collect these samples, researchers can combine data from existing studies. There are two main ways to combine existing data: through meta-analysis of summary statistics, and through Integrative Data Analysis using individual participant data.

**Meta-analysis**

This methodology is based on the summary statistics of groups provided in research reports. With a meta-analysis, researchers aim to find the summary effect size of their construct of interest. The summary effect size is represented by a weighted average of the effects included. Knowing the effect size across studies can help inform us about the true impact of a phenomenon. A second approach for meta-analysis is to understand the relations between multiple constructs. This can be done with meta-analytic structural equation modeling. Instead of focusing on single effect sizes, researchers compare and combine correlation matrices that are later used to estimate path analytic models.

*Steps*

1. Conduct a systematic review to find all published and unpublished reports that include the phenomenon or relations of interest
2. extract effect sizes and other relevant variables based on the summary statistics given in a report
3. statistically combine the effect sizes to obtain and estimate of the average effect size and confidence intervals or correlation matrix
4. if variation seems high, test for moderators of effect sizes
5. if of interest, conduct structural equation models to understand average relations between phenomena

***Additional reading on meta-analysis***

*Methodological*

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.

Cheung, M. W.-L., & Hafdahl, A. R. (2016). Special issue on meta-analytic structural equation modeling: Introduction from the guest editors. *Research Synthesis Methods*, *7*(2), 112–120. <https://doi.org/10.1002/jrsm.1212>

Gage, N. A., Cook, B. G., & Reichow, B. (2017). Publication bias in special education meta-analyses. *Exceptional Children*, *83*(4), 428–445.

Jackson, D., & Turner, R. (2017). Power analysis for random-effects meta-analysis. *Research Synthesis Methods*, *8*(3), 290–302. <https://doi.org/10.1002/jrsm.1240>

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178–206. <https://doi.org/10.3758/s13423-016-1221-4>

Schmidt, F. L., & Hunter, J. E. (2014). *Methods of meta-analysis: Correcting error and bias in research findings.* Sage publications.

Rosenthal, R. (1984). *Meta-Analytic Procedures for Social Science Research*. Sage Publication.

Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). How Many Studies Do You Need?: A Primer on Statistical Power for Meta-Analysis. *Journal of Educational and Behavioral Statistics*, *35*(2), 215–247. <https://doi.org/10.3102/1076998609346961>

*Examples from Psychology and Education*

Daucourt, M. C., Erbeli, F., Little, C. W., Haughbrook, R., & Hart, S. A. (2020). A meta-analytical review of the genetic and environmental correlations between reading and attention-deficit/hyperactivity disorder symptoms and reading and math. *Scientific Studies of Reading*, *24*(1), 23-56.

Joyner, R. E., & Wagner, R. K. (2020). Co-occurrence of reading disabilities and math disabilities: A meta-analysis. *Scientific Studies of Reading*, *24*(1), 14-22.

Roberts, G. J., Cho, E., Garwood, J. D., Goble, G. H., Robertson, T., & Hodges, A. (2020). Reading Interventions for Students with Reading and Behavioral Difficulties: A Meta-analysis and Evaluation of Co-occurring Difficulties. *Educational Psychology Review*, *32*(1), 17–47. <https://doi.org/10.1007/s10648-019-09485-1>

Suggate, S. P. (2016). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities*, *49*(1), 77–96.

Toste, J. R., Didion, L., Peng, P., Filderman, M. J., & McClelland, A. M. (2020). A Meta-Analytic Review of the Relations Between Motivation and Reading Achievement for K–12 Students. *Review of Educational Research*, *90*(3), 420–456. <https://doi.org/10.3102/0034654320919352>

**Integrative Data Analysis**

This methodology is based on raw, individual participant data. Unlike meta-analysis, in which the sample size is the number of studies included, IDA pools all individual data and creates one new data set. The goal of IDA is to create scaled scores on the constructs of interest across all independent data samples that will then be used as variables in subsequent statistical analysis. For each construct of interest, this is accomplished by selecting representative items from measures that are representative of this core construct and then modeling these items to create a valid and reliable scaled score. The items selected do not all have to be identical across the samples, and each sample can have unique items in addition to some items that are common across the samples. One method to create these scaled scores for IDA is using Moderated Nonlinear Factor Analysis (MNLFA). To estimate scaled scores across the independent data samples, MNLFA tests for measurement invariance across potential influential covariates at both the factor (intercept and variance) and item (intercept and variance) level. The following is an example of a path diagram for a MNLFA model on student behavior:



*Steps*

1. Acquire raw individual participant data from several projects
2. Check the accuracy of the data (for example, are there out of range values?)
3. Select items that correspond to the same construct
4. Harmonize items if needed
5. Estimate scaled score using MNLFA
6. Analyze the data according to your research question

**Additional reading on IDA:**

*Methodological*

Bauer, D. J., & Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. *Psychological Methods*, *14*(2), 101. <https://doi.org/10.1037/a0015583>

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, *14*(2), 81–100. <https://doi.org/10.1037/a0015914>

Curran, P. J., Hussong, A. M., Cai, L., Huang, W., Chassin, L., Sher, K. J., & Zucker, R. A. (2008). Pooling data from multiple longitudinal studies: The role of item response theory in integrative data analysis. *Developmental Psychology*, *44*(2), 365–380. <https://doi.org/10.1037/0012-1649.44.2.365>

Curran, P. J., McGinley, J. S., Bauer, D. J., Hussong, A. M., Burns, A., Chassin, L., Sher, K., & Zucker, R. (2014). A Moderated Nonlinear Factor Model for the Development of Commensurate Measures in Integrative Data Analysis. *Multivariate Behavioral Research*, *49*(3), 214–231. <https://doi.org/10.1080/00273171.2014.889594>

Hussong, A. M., Curran, P. J., & Bauer, D. J. (2013). Integrative Data Analysis in Clinical Psychology Research. *Annual Review of Clinical Psychology*, *9*(1), 61–89. <https://doi.org/10.1146/annurev-clinpsy-050212-185522>

*Examples from education*

Hornburg, C. B., Rieber, M. L., & McNeil, N. M. (2017). An integrative data analysis of gender differences in children’s understanding of mathematical equivalence. *Journal of Experimental Child Psychology*, *163*, 140–150. <https://doi.org/10.1016/j.jecp.2017.06.002>

Jansen, M., Lüdtke, O., & Robitzsch, A. (2020). Disentangling different sources of stability and change in students’ academic self-concepts: An integrative data analysis using the STARTS model. *Journal of Educational Psychology*, *advance online*.

Leijten, P., Raaijmakers, M., Wijngaards, L., Matthys, W., Menting, A., Hemink-van Putten, M., & Orobio de Castro, B. (2018). Understanding Who Benefits from Parenting Interventions for Children’s Conduct Problems: An Integrative Data Analysis. *Prevention Science*, *19*(4), 579–588. <https://doi.org/10.1007/s11121-018-0864-y>

**Benefits of IDA over meta-analysis:**

While potential more labor intensive, IDA has several advantages over meta-analysis.

*First*, because raw data is used, researchers are able to examine different relations within the data then those done by the original researchers.

*Second*, researchers using IDA can recheck the accuracy of the data and standardize the analyses across data sets. For example, if the original studies used two different ways to handle missing data, IDA provides the opportunity to reanalyze using only one of those.

*Third*, by combining individual data, researchers have opportunities to conduct subgroup analyses that were not included in the original studies, possible because combining data increases the number of individuals with low base-rate behaviors (such as low-incidence disabilities).

*Finally*, combining individual data increases the overall sample size and therefore provides the researcher with more power and the ability to estimate more complex models.

**Benefits of meta-analysis over IDA:**

The major advantage of meta-analysis over IDA is that it allows for a much broader sampling of the research conducted on a given topic. Not all data from all available studies will be made available. Meta-analysis provides the ability to conduct a more comprehensive research synthesis. It is also more flexible, in that the results obtained via IDA could be incorporated into a meta-analysis. That is, it would be possible to obtain datasets from studies that would inform a meta-analysis but the results from that project were not reported in a way to make them usuable for a meta-analysis. One could conduct an IDA analysis from those projects and incorporate the effect size estimates into a large meta-analysis that also includes the effects obtained from published studies.