SREE

2040 Sheridan Rd
Evanston, IL 60208

Society for Research on Educational Effectiveness
Advancing Education Research

www.sree.org
847.467.4001

Abstract Title Page

Title: A Re-Examination of the Education Production Function Using Individual Participant Data

Author(s): Therese D. Pigott
Ryan T. Williams
Joshua R. Polanin

Abstract Body

Background / Context:

A renewed effort to insure that publicly funded and collected data remains accessible to researchers has recently gained governmental and academic interests (Council on Governmental Relations, 2006). Organizations such as the Inter-University Consortium for Political and Social Research (ICPSR) have long archived and collected large data sets, and the National Institute of Health (NIH), and the National Science Foundation (NSF) both have formal requirements for grantees around plans for sharing and archiving data. These databases have the potential to enable advanced analysis for both policy and practice.

While these data archives provide researchers the opportunity to perform secondary analyses, they also engender the opportunity for new methods of meta-analysis. In medicine, where individual patient data is more commonly available than in the social sciences, methodologists have outlined a number of methods for combining individual participant data with the more traditional aggregated data usually collected in a meta-analysis. The purpose of this presentation is to illustrate methods of meta-analysis that combine both individual participant data (IPD) and aggregated data (AD) from traditional meta-analyses. Our example is based on an on-going project that uses data from Greenwald, Hedges, and Laine's (1996) meta-analysis of 60 primary research studies that synthesized aggregated data on education production functions. At least six of the studies included in this meta-analysis used data from publicly available data sets. The presentation will compare the results from traditional aggregated data meta-analysis with a range of methods that incorporate both aggregated and individual level data.

Cooper & Patall (2009) recently outlined the benefits and limitations of IPD meta-analysis for issues in the social sciences. The advantages of incorporating individual participant data include but are not limited to:

- Increased collaboration across researchers: As mentioned earlier, the National Science Foundation and the National Institutes of Health both have developed policies for data sharing. The National Institutes of Health (2003) statement on sharing research data indicates that all applications with direct costs above $500,000 must address data sharing. Curran & Hussong (2009) and Shrout (2009) both provide examples of collaborations that have been developed around the pooled data sets.
- Obtaining missing data and checking original analyses: One advantage Cooper & Patall (2009) cite for IPD is the ability to check the original data from the primary studies, and to fit models that were not possible with only the data provided in the studies. For example, the primary data set may include outcome measures or characteristics of participants not reported in the original study. The problem of outcome reporting bias has been discussed by Orwin & Cordray (1985) in the social science literature, and is a source of considerable discussion in medicine (Turner, Matthews, Linardatos, Tell, & Rosenthal, 2008; Vedula, Bero, Scherer, & Dickersin, 2009). Missing data is also a problem in aggregated data analysis (Pigott, 2009) when particular moderators of effect size are not all reported across studies or when information to compute an effect size is not present. With the original data, effect sizes can be computed with full information, and analyses of effect size variation can use more detailed background characteristics of the study and participants.

- Increased statistical power: Another potential benefit of IPD meta-analysis is in statistical power. Simmonds & Higgins (2007) compare the power for detecting interactions among study level characteristics and effect sizes in an AD meta-analysis versus an IPD meta-analysis. Under many conditions, an IPD meta-analysis has greater power than an AD meta-analysis. These conditions depend on the variation in the variables that are potential moderators of study effect both within and between studies.
- Examining differential effects while minimizing aggregation bias: As researchers developing multi-level modeling have long stressed, aggregation bias operates within nested educational data (Raudenbush & Bryk, 2002) and should be carefully monitored in conclusions of AD meta-analysis (Cooper & Patall, 2009; Schmid et al., 2004). Having the individual participant data allows the examination of differences in treatment effectiveness at the level of the individual rather than at the level of the study. Being able to make inferences at the individual participant level not only avoids aggregation bias, it may lead to inferences for a meta-analysis that are more readily applied to practice.
- Broadening the psychometric evaluation of constructs: As illustrated by Bauer & Hussong (2009), pooled individual participant data provides the opportunity to examine the psychometric properties of measures used across studies, and in some cases to develop commensurate measures across studies. In AD meta-analysis, by necessity we assume that measures of a similar construct are comparable, but without the individual level data and information about the measures' psychometric properties, we cannot be sure if these measures share similar properties. Under certain conditions, say when studies share some items in common, measures of a construct could be linked across studies using item response theory. Comparing across measures may also lead to the development of more sensitive assessments that can be shared across studies.
- Allowing more complex analyses of the primary data: Much of the research on meta-analysis methodology in the social sciences focuses on methods for combining results of complex statistical analyses across studies. For example, many reviewers are forced to exclude studies that report regression results since we do not have methods for combining across different regression models. With individual level data, problems with combining across different regression models could be alleviated by estimating similar models across studies with the original data.

Purpose / Objective / Research Question / Focus of Study:

The focus and purpose of this research is to examine the benefits, limitations, and implications of Individual Participant Data (IPD) meta-analysis in education. Comprehensive research reviews in education have been limited to the use of aggregated data (AD) meta-analysis, techniques based on quantitatively combining information from studies on the same topic. These analyses have obvious benefits, but can at times be limiting.

The proposed project will conduct an IPD meta-analysis on studies focused on estimating an education production function. Our research goal is to understand the benefits and limitations of IPD meta-analysis compared to AD meta-analysis. More specifically, our research questions are the following:

1. What are the methods suggested in the literature for conducting a meta-analysis that combines both aggregated data and individual participant data?
2. How do the results of the original analysis compare with the meta-analysis using both aggregated and individual data?

3. What are potential benefits and limitations of the use of IPD meta-analysis in the social sciences?

Significance / Novelty of study:

Although medical researchers have conducted IPD meta-analyses, little attention has been paid to its use in education or psychology (Curran & Hussong, 2009). Cooper & Patall (2009) discussed the benefits and limitations in general for IPD versus AD meta-analysis, but did not conduct an analysis. In psychology, a special issue of *Psychological Methods* is devoted to issues related to Integrated Data Analysis (Bauer & Hussong, 2009; Cooper & Patall, 2009; Curran, 2009; Curran & Hussong, 2009; Hofer & Piccinin, 2009; Shrout, 2009). To date, the only meta-analysis in education utilizing individual-level data was conducted by Goldstein, Yang, Omar, Turner & Thompson (2000) combining data from studies of the effects of class size with primary data from the Tennessee STAR experiment. This study will utilize 3-4 datasets initially utilized in studies that are included in the Greenwald, Hedges, and Laine (1996) meta-analysis to compare the findings of these two methods and assess the feasibility and cost of an IPD analysis.

Statistical, Measurement, or Econometric Model:

The data analysis will compare two strategies for analyzing a mix of individual level and aggregated level data. The one-stage method is based on multi-level modeling techniques, while the two-stage method first obtains the aggregated data for all studies (either from IPD or from study reports), and then synthesizes the results. In order to present the models that will be used, we begin with a brief outline of random effects models for aggregated data meta-analysis, and then develop the models that will be used in this project.

Random effects model of effect size with aggregated data. The typical random effects model of effect size can be written using a two-level hierarchical model as outlined by Raudenbush (2009). We compute an effect size T_i from study *i,* where *i*= 1,…, *k*, and also the variance of that effect size, v_i, using statistics for that study. Level 1 is given as

$$T_i = \theta_i + e_i,\ e_i \sim \eta(0, v_i) \qquad 1$$

In AD meta-analysis, we assume that v_i is known. Level 2 is given as

$$\theta_i = \theta + u_i,\ u_i \sim \eta(0, \sigma_\theta^2), \qquad 2$$

where θ is the overall mean effect size, and the variance component is given as σ_θ^2. The random effects variance can be estimated either directly using the method of moments or with restricted maximum likelihood as suggested by Raudenbush (2009).

Random effects model for IPD. In order to illustrate the one- and two-stage methods, we need a model for the data from a study that provides individual participant data. The outcome for IPD is the individual participant's response on the target measurement, denoted for participant *j,* in study *i,* as y_{ij}. Note that in AD meta-analysis, we use summary statistics from the primary reports of research to compute an effect size for each study. In order to make the outcomes parallel in an IPD analysis with an AD analysis, we will use the standardized outcome, denoted here by y_{ij} for student $j, j = 1,\ldots, n_j$, in study *i,* *i*=1,…,*k.* Thus, each students' outcome will be standardized using the overall mean and standard deviation of the outcome observed in that study. We do this so that we can synthesize study outcomes that are not using the same measure of a construct. We can write a hierarchical linear or mixed model for our IPD data following Riley et al. (2008). The model is given as

$$
\begin{aligned}
y_{ij} &= \phi_i + \theta_i x_{ij} + e_{ij} \\
\theta_i &= \theta + u_i \\
u_i &\sim \eta(0, \sigma_\theta^2) \\
e_{ij} &\sim \eta(0, 1)
\end{aligned}
\qquad 3
$$

where x_{ij} is a 0/1 code designating control or treatment group membership, the fixed study effect is ϕ_i, with the random treatment effect in study *i* given by θ_i. The variance within each study for the outcome is 1, since we have standardized our outcomes, and the variance for the θ_i is σ_θ^2. Our goal in an IPD meta-analysis would be to estimate the mean treatment effect, θ, and its variance, σ_θ^2, using standard methods of hierarchical linear models (Raudenbush & Bryk, 2002).

Two-stage method with both IPD and AD. As Riley et al. (2008) find, the easiest method to employ with a mix of IPD and AD is a two-stage model. The researcher first computes the study level effect sizes from each IPD study, and then continues with estimating the random effects model given in Eqn. 1.

One-stage method with both IPD and AD. One-stage methods for a mix of IPD and AD would be analogous to fitting hierarchical linear models when some of the level-2 units do not provide individual level data. In Riley et al.'s formulation, we assume that AD studies provide their effect size estimate, $T_i = \hat{\theta}_i$, and its variance v_i, which is known. For studies that contribute individual level data, our model is given in Eqn. 5 where y_{ij} is our standardized outcome for student *j* in study *i*. In the combined IPD and AD model, studies with IPD contribute individual student outcome data, while the AD studies are assumed to have only one student with an outcome equal to $\hat{\theta}_i$ and residual variance known and equal to v_i, the variance of the particular effect size used in the analysis. Following Goldstein et al. (2000), Riley et al. add a dummy code, D_i, that takes the value 1 for IPD studies, and 0 for AD studies. The dummy code allows the estimation of the treatment effect $\hat{\theta}_i$ for studies with IPD, and both AD and IPD studies to contribute to the estimation of the average treatment effect, θ, and the between-study variance in the treatment effect, σ_θ^2. Riley et al. give the model as

$$
\begin{aligned}
y_{ij}^* &= D_i \phi_i + \theta_i x_{ij} + e_{ij}^* \\
\theta_i &= \theta + u_i \\
u_i &\sim \eta(0, \sigma_\theta^2) \\
e_{ij}^* &\sim \eta(0, v_i^*)
\end{aligned}
\qquad 4
$$

As stated above, for each IPD study, the outcome is $y_{ij}^* = y_{ij}$ and the $v_i^* = 1$ since we have standardized our outcomes within studies. In the AD studies, we assume only one observation (*j*=1), and we set x_{i1}=1. The response in the AD studies is the estimate of the effect size in that study, $y_{i1}^* = \hat{\theta}_i$, and variance v_i.

Usefulness / Applicability of Method:

The usefulness and applicability of this method is dependent on the availability of primary data. A traditional AD can produce results strictly from primary studies, the usefulness and power of an IPD requires a number of datasets. Although this can be a limitation, we believe the call to action set forth by the NSF (among others) to disseminate primary data and the availability of communication ease (i.e. the internet) will only increase the applicability of these methods.

Data Collection and Analysis:

Original findings, as mentioned previously, derive from Greenwald, Hedges, and Laine (1996). GHL synthesized 60 primary studies, of which 36 utilized a large-scale dataset. Unfortunately many of these studies were conducted prior to the use of computers (many were conducted in the 1970's and 1980's) and therefore data are not available. However, we have obtained three datasets and in the process of acquiring a fourth. These include:

- Equal Education Opportunity Survey
- Project Talent
- High School and Beyond
- Working to obtain a dataset from Illinois, Kentucky, and California

All datasets include a measure of student achievement and at least one measure of (or proxy to) per-pupil expenditure.

Conclusions:

The methods proposed have the potential to further meta-analytic techniques. Although the "traditional" aggregated data analysis will remain paramount, the IPD analysis will engender and enable more sophisticated and precise estimates of treatment effects. Of course, data sharing limitations, data accessibility, and time restrictions potentially limit the capabilities of IPD, the ability to test multiple variables, the elimination of aggregation bias, and increased statistical power embolden IPD's prowess. Indeed IPD may be the future of meta-analysis and our proposal provides the opportunity to further the method's research.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Bauer, D. J., & Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. Psychological Methods, 14(2), 108-125.

Cooper, H., & Patall, E. A. (2009). The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. Psychological Methods, 14, 165-176.

Curran, P. J. (2009). The seemingly quixotic pursuit of a cumulative psychological science: Introduction to the special issue. Psychological Methods, 14(2), 77-80.

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. Psychological Methods, 14(2), 81-100.

Goldstein, H., Yang, M., Omar, R., Turner, R., & Thompson, S. (2000). Meta-analysis using multilevel models with an application to the study of class size effects. Applied Statistics, 49, 399-412.

Greenwald R., Hedges, L.V., & Laine, R.D. (1996). The effect of school resources on student achievement. *Review of Educational Research, Vol 66*(3), p. 361-396.

Hofer, S. M., & Piccinin, A. M. (2009). Integrative data analysis through coordination of measurement and analysis protocol across independent longitudinal studies. Psychological Methods, 14(2), 150-164.

Orwin, R. G., & Cordray, D. S. (1985). Effects of deficient reporting on meta-analysis: A conceptual framework and reanalysis. Psychological Bulletin, 97(1), 134-147.

Pigott, T. D. (2009). Handling missing data. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), The handbook of research synthesis and meta-analysis (2nd ed., pp. 399-416). New York: Russell Sage Foundation.

Raudenbush, S. R., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W. (2009). Analyzing effect sizes: Random effects models. In H. Cooper, L.V. Hedges, & J. C. Valentine, *The handbook of research synthesis and meta-analysis.* New York: Russell Sage Foundation.

Schmid, C. H., Stark, P. C., Berlin, J. A., Landais, P., & Lau, J. (2004). Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors.

Simmonds, M. C., & Higgins, J. P. T. (2007). Covariate heterogeneity in meta-analysis: Criteria for deciding between meta-regression and individual patient data. Statistics in Medicine, 26, 2982-2999.

Shrout, P. E. (2009). Short and long views of integrative data analysis: Comments on contributions to the special issue. Psychological Methods, 14(2), 177-181.

Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A., & Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. The New England Journal of Medicine, 358, 252-260.

Vedula, S. S., Bero, L., Scherer, R. W., & Dickersin, K. (2009). Outcome reporting in industry-sponsored trials of gabapentin for off-label use. The New England Journal of Medicine, 361, 1963-1971.