Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 8, Number 15, June, 2003 ISSN=1531-7714

Meta-Analysis Methods for Synthesizing Treatment Effects in Multisite Studies: Hierarchical Linear Modeling (HLM) Perspective

Sema A. Kalaian
University of Toledo

A multisite study can be defined as the conduct of a research to study the effectiveness of similar or variations of the same interventions across multiple similar or distinct sites. These sites can involve multiple clusters of individuals, clinics, schools, classrooms, geographic locations, contexts, or field settings. In the past several years, multiple site research has been identified in the social, behavioral, and health literature as cluster evaluation, multisite research, cross-site evaluation, multi-center clinical trials, or multiple field research.

The popularity of multisite interventions in health, medicine, social and behavioral disciplines can be attributed to its many advantages and desirable features. First, it is much easier and quicker to recruit a sample of individuals in a short period of time for a multisite intervention than for a large geographically scattered national sample. Second, it is much less expensive to implement the intervention on a sample of individuals who are geographically clustered than to individuals who are widely dispersed across the nation. Third, the research process is easier to manage for individual sites because the project/research staff at each site follows the same research principles, trial procedures, clinical protocols, and guidelines. Fourth, multisite interventions improve the representativeness of the study population. Finally, multisite interventions increase the generalizability and validity of the research conclusions across multiple similar or distinct field settings and contexts (Raudenbush & Liu, 2000).

Although meta-analysis methods for synthesizing treatment effects (effect sizes) from multiple primary studies have been utilized in almost all social, medical, psychological, and behavioral disciplines, methods of applying meta-analytic techniques to synthesize research findings across large-scale multisite studies have been sparse. For instance, Schafer (2001) described two multiple field replications research examples that applied fixed-effects meta-analytic methods to synthesize the findings and enhance the evidence available regarding the effectiveness of the interventions. The first multiple field research example was studied by Guthrie, Schafer, Von Secker, and Alban (2000) to synthesize the relationships between instructional characteristics of schools and the degrees of gain or loss in student achievement over two years. The second example was studied by Schafer, Swanson, Bené, and Newberry (2001) to examine the effects of teacher knowledge of rubrics as an innovative instructional method on high school student achievement. Also, Banks, & et al. (2002) described and used a prospective meta-analytic approach to analyze a multisite study of homeless prevention. The present study extends Schafer's (2001) and Banks, & et al. (2002) work and presents a routine and flexible statistical mixed-effects meta-analytic methodology that can be used by multisite evaluators from diverse fields (e.g. education, psychology, health, and social research) who are interested in the effects of innovative interventions across multiple sites. Contrary to fixed-effects meta-analysis approach, this mixed-effects meta-analytic methodology conceptualizes the multiple sites as being a random sample from a universe or population of multiple sites. Thus, the proposed methodology will lead to (a) increased generalizability of multisite results to a population of sites not included in the original multisite study, and (b) draw rich and sound conclusions about program effectiveness in individual sites as well as across sites.

The objectives of the present mixed-effects meta-analytic application are to provide practical guidelines to: (a) Calculate treatment effect sizes from multiple sites; (b) Calculate the overall mean of the site effect sizes and their variances; (c) Model the heterogeneity in these site treatment effects as a function of site and program characteristics plus unexplained random error using Hierarchical Linear Modeling (HLM); (d) Improve the ability of multisite evaluators and policy makers to reach sound conclusions about the effectiveness of educational and social interventions based on multisite evaluations; and (e) Illustrate the proposed methodology by applying these methods to real multi-site research data.

Multisite Illustrative Example

The proposed mixed-effects meta-analytic methodology via Hierarchical Linear Modeling is an extension to the prospective fixed-effects meta-analytic approach proposed by Schafer (2001) and Banks et al. (2002). This proposed methodology is applied to a national multisite evaluation study to assess the effects of new innovative head start like services across 48 sites. The study is a national quasi-experimental study with treatment and control groups in each site. One of the key purposes of this national evaluation study was to determine the effectiveness of comprehensive head

start like services for kindergarten children in improving student reading and mathematics achievement. The proposed meta-analytic methodology will help us to answer the following key research questions: In which sites the innovative program was effective? Was the innovative program effective in increasing math performance? And, what are the site level factors that contributed to the program effectiveness?

Although the evaluation study is real, the real data and identity of the study will not be revealed here because the purpose of this application is to illustrate the application of the proposed meta-analytic methodology and not to draw any conclusions about the effectiveness of the intervention in this national multisite evaluation study.

Results of applying the Mixed-Effects Meta-Analysis

The present study's analytic method involves the conceptualization and application of the mixed-effects model via Hierarchical Linear Modeling (Kalaian, 1994; Kalaian & Raudenbush, 1996; Raudenbush & Bryk, 2002) for meta-analysis to the above mentioned large-scale multisite evaluation study. Meta-analytic techniques are applicable when the meta-analyst faces the challenge of summarizing independent empirical primary studies that compare an experimental group to a control group on a key outcome measure (Hedges & Olkin, 1985; Kalaian & Raudenbush, 1996). In this study, the statistical analysis results from different sites are treated as a collection of related primary studies in this meta-analysis application. Accordingly, each site will be considered as a primary study.

The application of this proposed methodology to multisite studies can be carried on using the following three integrated general meta-analytic steps. For the sake of continuity of the presentation, the statistical formulas are provided to carry on these three general steps.

Step I: Calculating Site Effect Sizes

The emphasis of this study is on the application of mixed-effects meta-analytic methods via Hierarchical Linear Modeling to randomized experimental/clinical trials research studies with *K* sites, each comparing an outcome measure of clinical/experimental treatment (*E*) to the same outcome of the control condition (*C*) in site *i* (*i=1,2,…K*).

The first step in the proposed meta-analytic application was the calculation of effect sizes for each of the individual sites. Thus, the index to measure the effectiveness of an experimental treatment in this multisite study is based on the standardized mean effect-size outlined by Raudenbush and Bryk (2002) and can be computed for an outcome measure in each individual site. For instance, a site with an outcome measure (e.g., math achievement scores) for an experimental and a control group, would yield one effect size for each site.

The estimated standardized mean difference (Hedges, 1994) between experimental and control groups for the outcome measure, in the *i*th site is

$$g_i = \frac{\overline{Y}_i^E - \overline{Y}_i^C}{S_i} \qquad (1)$$

where $\overline{Y}_i^E$ and $\overline{Y}_i^C$ are the i^{th} experimental and control group means respectively. Also, S^2_i is the pooled within-groups estimate of the sample variances and it is calculated as follows

$$S_i^2 = \frac{\sum_p (n_i^E - 1)(S_i^E)^2 + \sum_p (n_i^C - 1)(S_i^C)^2}{\sum_p (n_i^E - 1) + \sum_p (n_i^C - 1)} \qquad (2)$$

where S^E_i and S^C_i are the experimental and control group standard deviations for the outcome measure in site *i* respectively.

If the sample sizes within each site are small, the g_i is an upwardly biased estimator of the population effect size δ_i and can be corrected with the following formula

$$d_i = \left(1 - \frac{3}{4m_i - 1}\right) g_i \qquad m_i = n^E_i + n^C_i - 2. \qquad (3)$$

Across the 48 sites, the values of math treatment effect sizes ranged from -0.36 to 1.11. Although more than 50% of these effect sizes were positive, the magnitudes of these effects seemed to be quite variable.

Step II: Calculating the Variances of Site Effect Sizes

The second step in the proposed meta-analytic application was the calculation of the variances of the calculated site effect sizes from the first step. The variances of these site effect sizes can be calculated using the following formula

$$\hat{\sigma}^2(d_i) = \frac{n_i^E + n_i^C}{n_i^E n_i^C} + \frac{d_i^2}{2(n_i^E + n_i^C)}. \quad (4)$$

Step III: Modeling Site Effect Sizes

The third step in the proposed meta-analytic methodology is the application of the mixed-effects linear model via Hierarchical Linear Modeling (Kalaian & Raudenbush, 1996; Raudenbush & Bryk, 2002) to these computed site effect sizes and their variances. Presentation of this model in two stages clarifies the logic of the hierarchical structure of the meta-analytic data. At the first stage, a "within-site" model specifies the site effect size as a function of the true site effect size and sampling error. The second stage, a "between-sites" model, specifies the distribution of the true site effect sizes from the first stage as a function of study characteristics and random errors (Raudenbush & Bryk, 2002).

Thus, practically speaking, in order to analyze any multisite data using the proposed methodology, two data files are needed to use the Hierarchical Linear Modeling (HLM) software (Raudenbush, Bryk, Cheong, & Congdon, 2000). One data file includes the site identification numbers, site effect-sizes, and their variances. The second data file consists of the site identification numbers and coded site characteristics.

Within-site model

In the within-site model for the mixed-effects meta-analysis, the calculated site effect size, d_i, of site i, depends upon a population site effect size δ_i plus a random sampling error, e_i . Thus, the basic within-site model for site i can be represented as

$$d_i = \delta_i + e_i , \qquad i = 1,2,\ldots,K. \quad (5)$$

Between-Sites Model

At the second HLM modeling stage, the between-site model can be expressed in two forms and can be used simultaneously for different meta-analysis purposes. The first is the unconditional model, where we assume that the site effect-size parameter, δ_i from the within-site model, varies around a grand mean plus a random error. The results of applying the unconditional model helps the researcher in assessing the overall weighted site effect-size average and examine the homogeneity of the site effect sizes. The second is the conditional model, where we assume that the effect-size parameter d_i depends on known site and program characteristics plus random error and this model helps the evaluator to explain the variations among the site effect sizes.

Unconditional Between-Sites Model (Assessing the Heterogeneity of Site Effect Sizes). This model is the first and the basis for several subsequent models that form the logical hierarchy in this proposed meta-analytic application to multisite studies. This model helps to (1) calculate the mean effect sizes across sites; (2) test the hypothesis that the mean effect size equals zero; and (3) examine and test the heterogeneity of site effect sizes across sites.

In this unconditional model, where no predictor variables (site characteristics) are modeled, the true effect size parameters are viewed as varying randomly around an overall grand mean, γ, plus random error U_i

$$\delta_i = \gamma + U_i , \qquad U_i \sim N(0 , \tau) . \quad (6)$$

Here τ represents the amount of variation in the site effect sizes. The results of applying this unconditional model suggest an overall weighted average treatment effect of 0.03 and significant heterogeneity in these effect sizes (τ_{math} = 0.013 with 46 degrees of freedom, p= .001). These results indicate that a single population site effect size does not underlie the data.

Since the results of the initial unconditional analysis has significant χ^2 value, which indicates that significant heterogeneity (inconsistency) exists among the site effect sizes, interest turns to establishing and modeling a conditional between-sites model.

Conditional Between-Sites Model (Modeling the Heterogeneity of Site Effect Sizes). In this model, which can be considered an expansion of the unconditional model, we use information about site characteristics (site program,

treatment conditions, and subject's characteristics) to account for the variation among the effect sizes. In other words, we try to explain the variations in the effect size parameters by knowing program level factors, contextual, and treatment variations in the multiple sites under consideration. This between-study model can be written in the following form

$$\delta_i = W_i \Upsilon + U_i , \quad U_i \sim N(0 , \tau) \qquad (7)$$

where, δ_i and U_i are the true effect size and the associated random error for each site. W_i is an m by q matrix of known study characteristics and Υ is a (q x 1) vector of between-site parameters. Here, τ is the amount of unexplained parameter variation left after knowing the effects of known site characteristics.

The finding of significant heterogeneity values in the previous no-predictor model (unconditional model), led to fit a series of specific HLM conditional models to this data, including potential predictor variables (e.g., Head Start program duration and intensity) to account for the sources of variability in these site effect sizes. As shown in Table 1, the results of this analysis show that program duration contributed significantly in explaining the variation in math effect sizes ($\beta_{duration}$ coefficient = 0.113, p = .05).

Table 1: HLM Conditional Modeling of Variation in Math Achievement Effect Sizes across Sites

Variables in the model	Coefficient	P-value	% Variance explained	Unexplained random variation
Intercept	-0.08	0.088		
Program intensity	-0.09	0.090	4.7%	0.01*
Program duration	0.04	0.052		

Conclusion and Discussion

One of the major purposes of multisite research is to compare the effects of an experimental treatment across multi sites to determine cross-site similarities and disparities. This can be accomplished by estimating individual site's intervention effect size. The other key purpose is to combine these site effect sizes across sites, as well as taking into account site related characteristics using meta-analytic statistical methods. By applying the proposed methodology to multisite evaluation studies, the global statistical summary results and conclusions about program effectiveness for individual sites and across sites can be made more meaningful and accurate for multisite researchers and policy makers.

Given the fact that multiple site and cluster evaluations represent an opportunity to increase the generalizability of evaluation findings, the proposed statistical methodology will have a potential to improve knowledge about the effectiveness of intervention alternatives in most disciplines that deal with multi-site research designs. Disciplines such as social science, education, medicine, health sciences, and others will benefit from this statistical methodology application to be able to better investigate and have sound and defensible conclusions about program effectiveness for individual sites. Thus, this study could serve as a step by step primer for future consideration of mixed-effects meta-analysis techniques as statistical tools for multisite studies.

As demonstrated, the application of the meta-analytic via Hierarchical Linear modeling to multisite studies is quite flexible and useful for drawing accurate and comprehensive conclusions about program effectiveness across sites as well as for individual sites. Thus, this HLM meta-analytic application to multisite studies intended to yield complementary and supplementary methodology for multisite external inferences that combine the flexibility of model-based inferences with the reliability and generalizability of multisite design based inferences.

References

Banks, S., McHugo, G. J., Williams, V., Drake, R. E., & Shinn, M. (2002). A Prospective Meta-Analytic Approach in a multisite Study of Homelessness Prevention. *New Directions for Evaluation,* 94, 45-59.

Becker, B. J. (2000). Multivariate Meta-Analysis. In H. E. A. Tinsley and S. Brown (Eds). *Handbook of Applied Multivariate Statistics and Mathematical Modeling*. Academic Press. San Diego.

Guthrie, J. T., Schafer, W. D., Von Secker, C., & Alban, T. (2000). Contributions of instructional practices to reading

achievement in a statewide improvement . *Journal of Educational Research,* 93, 211-225.

Hedges, L.V., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. New York: Academic Press.

Hedges, L.V. (1994). Fixed Effects Models. In H. Cooper and L. V. Hedges (Eds). *The Handbook of Research Synthesis*. New York: Russell Sage Foundation.

Kalaian H. A. (1994). *A Multivariate Mixed Linear Model for Meta-Analysis*. Unpublished Doctoral Dissertation, Michigan State University, East Lansing.

Kalaian, H. A., & Raudenbush, S. W. (1996). A multivariate Linear Model for Meta-Analysis. *Psychological Methods*, 1(3), 227-235.

Raudenbush, S. W., & Liu, X. (2000). Statistical Power and Optimal Design for Multisite Randomized Trials. *Psychological Methods*, 5(2), 199-213.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publications.

Raudenbush, S. W., & Bryk, A. S., Cheong, Y., & Congdon, R. T. (2000). *HLM 5: Hierarchical linear and nonlinear modeling*. Chicago: Scientific Software International.

Schafer, W. D. (2001). Replication: a design principle for field research. *Practical Assessment, Research & Evaluation,* 7(15). Available online: http://pareonline.net/getvn.asp?v=7&n=15

Schafer, W. D., Swanson, G., Bené, N., & Newberry, G. (2001). Effects of teacher knowledge of rubrics on student achievement in four content areas. *Applied Measurement in Education,* 14, 151-170.

Descriptors: Meta Analysis; Research Methodology; Research Reports; Synthesis; Causal Models; HLM; Predictor Variables; Research Design; Statistical Inference

Citation: Kalaian, Sema A. (2003). Meta-analysis methods for synthesizing treatment effects in multisite studies: hierarchical linear modeling (hlm) perspective. *Practical Assessment, Research & Evaluation*, 8(15). Available online: http://PAREonline.net/getvn.asp?v=8&n=15.