



REFLECTIONS ON METHODOLOGY

APRIL 2018

NEW YORK

16 East 34th Street
New York, NY 10016
Tel: 212 532 3200

OAKLAND

475 14th Street
Suite 750
Oakland, CA 94612
Tel: 510 663 6372

WASHINGTON, DC

1990 M Street, NW
Suite 340
Washington, DC 20036

LOS ANGELES

11965 Venice Boulevard
Suite 402
Los Angeles, CA 90066

www.mdrc.org



Quantifying Cross-Site Impact Variation

SOME IMPORTANT LESSONS

By *Howard S. Bloom and Michael J. Weiss*

This post is one in a series highlighting MDRC's methodological work. Contributors discuss the refinement and practical use of research methods being employed across our organization.

In a recently published three-article set, MDRC researchers and colleagues discuss quantifying cross-site impact variation using data from multisite randomized trials. The papers show that the extent to which the effects of an intervention vary across settings has important implications for policy, practice, science, and research design. This post distills some key considerations for research design and for reporting and interpreting cross-site impact variation.

The three articles, published in the *Journal of Research on Educational Effectiveness*, are based on research funded by the Spencer Foundation, the William T. Grant Foundation, and the Institute for Education Sciences. The first paper ([Bloom et al., 2017](#)) considers how to estimate, report, and interpret cross-site impact variation.¹ The second paper ([Bloom and Spybrook, 2017](#)) considers how to design multisite trials with adequate precision in the presence of this variation. The third paper ([Weiss et al., 2017](#)) applies methods from the first two papers to data from 16 multisite trials in education and training research to quantify the cross-site impact variation they reflect.

The benefits of understanding variation apply on multiple levels. Local policymakers and practitioners need to know both the average impact of an intervention and its variation across settings to properly assess its likely benefits and risks for their jurisdictions. For social scientists, cross-site impact variation offers opportunities to learn about mechanisms or mediators through which interventions produce their impacts and characteristics of settings and sample members that influence or moderate these impacts. And for researchers designing studies, cross-site impact variation can markedly affect the statistical precision of effect estimates and hence influence the sample size requirements for these estimates.

One important lesson illustrated by our papers involves **reporting cross-site impact variation** and reflects the difference between variation in impact estimates and variation in true impacts. It is a simple matter to produce internally valid estimates of the mean impact of an intervention for each site in a multisite randomized trial. However, reporting cross-site variation in these estimates through a frequency distribution or a standard deviation can greatly overstate the amount of true variation that exists. This can occur because differences between site-specific impact estimates have two sources: (1) true cross-site impact differences and (2) differences in random, site-specific estimation error. For studies without very large site samples, most of the variation in site-specific impact estimates reflects random estimation error. Thus, it is essential to use a rigorous method for inferring the magnitude of true cross-site impact variation.

Another lesson involves **interpreting cross-site impact variation** and reflects the fact that the impact of an intervention is by necessity defined with respect to a specific counterfactual alternative or set of alternatives. As [Holland \(1986, p. 950\)](#) aptly notes, “the effect of a cause is always relative to another cause.” For example, the impact of lottery-based assignment to a charter school is

¹This paper received the 2017 Outstanding Article award from the *Journal of Research on Educational Effectiveness*.

defined as the difference between the mean outcome for students assigned to the charter school and the mean outcome for comparable students not assigned to it and thus attending other schools. Consequently, cross-site impact variation may reflect the fact that both (1) charter schools can vary in their ability to produce educational gains for students with a given educational background and potential, and (2) alternatives to charter schools can also vary in this regard. Indeed it is possible, in principle, to have no such variation in the effectiveness of specific charter schools but considerable variation in the corresponding effectiveness of their counterfactual schools, and thus to have considerable variation in charter school *impacts*. This very real fact of evaluation life illustrates the importance of focusing not just on the treatment being studied but also on the treatment *contrast*.

Interpretation of cross-site impact variation is also complicated by the possibility that some observed variation in impacts between sites is due to differences in the composition of their sample members; likewise, some observed variation in impacts between subgroups of individuals may be due to differences in their distribution across sites. Thus when studying these sources of impact variation it is essential to account for their potential conflation.

Once you acknowledge the possibility of impact variation across sites or individuals, it is essential to **clearly specify your target population**. Researchers must decide whether to limit their estimates of a mean intervention impact to the sites in their sample (a fixed-effect inference) or to project those estimates to a superpopulation of sites represented by their sample (a random-effects inference).² In addition, researchers must decide whether to infer study findings to a population of sites or a population of individuals. Specifying these aspects of your target population has important implications both conceptually (it helps to define your parameters of interest) and for estimation (it determines how sites and sample members should be weighted).

Yet another lesson from our research is the importance of specifying — and where possible assessing — the **full range of assumptions** that underlie the model used to estimate cross-site impact variation. In this regard, seemingly little things (like specifying the individual-level error distribution) can make a big difference in the magnitude and statistical significance of estimates of cross-site impact variation.

Finally, it is important to note that **there may be no clear relationship between cross-site impact variation and the magnitude of overall mean impact**. The present empirical research identified interventions with (1) near-zero mean impact and substantial cross-site impact variation (charter schools in multiple states); (2) substantial mean impact and substantial impact variation (New York City's small high schools of choice); (3) near-zero mean impact and near-zero impact variation (after-school reading programs in multiple states); and (4) substantial mean impact and near-zero impact variation (high-school career academies in multiple states). These diverse findings illustrate the need for future empirical research on the patterns of impact variation that exist and factors that predict these patterns.

The presence of impact variation across sites and individuals offers a rich opportunity to learn how interventions work, if at all, and for whom. Careful attention to the complications involved is crucial to ensure the validity of the analysis.

²This broader inference can be justified by the fact that even for a convenience sample of sites (the basis for most past multisite trials) the ultimate target of interest is typically not just sites in a study's sample but rather some population of sites represented by the sample where the intervention being tested might be implemented.

REFERENCES

Bloom, Howard S., Stephen W. Raudenbush, Michael J. Weiss, and Kristin Porter. 2017. "Using Multi-site Experiments to Study Cross-Site Variation in Treatment Effects: A Hybrid Approach with Fixed Intercepts and a Random Treatment Coefficient." *Journal of Research on Educational Effectiveness* 10, 4: 817-842. <https://doi.org/10.1080/19345747.2016.1264518>.

Bloom, Howard S., and Jessaca Spybrook. 2017. "Assessing the Precision of Multisite Trials for Estimating the Parameters of a Cross-Site Population Distribution of Program Effects." *Journal of Research on Educational Effectiveness* 10, 4: 877-902. <https://doi.org/10.1080/19345747.2016.1271069>.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81, 396: 945-960. <https://doi.org/10.1080/01621459.1986.10478354>.

Weiss, Michael J., Howard S. Bloom, Natalya Verbitsky-Savitz, Himani Gupta, Alma E. Vigil, and Daniel N. Cullinan. 2017. "How Much Do the Effects of Education and Training Programs Vary Across Sites? Evidence from Past Multisite Randomized Trials." *Journal of Research on Educational Effectiveness* 10, 4: 843-876. <https://doi.org/10.1080/19345747.2017.1300719>.