

## Using a Multi-Site RCT to Predict Impacts for a Single Site: Do Better Data and Methods Yield More Accurate Predictions?

Robert B. Olsen, Larry L. Orr, Stephen H. Bell, Elizabeth Petraglia, Elena Badillo-Goicoechea, Atsushi Miyaoka, Elizabeth A. Stuart

Multi-site randomized controlled trials (RCTs) produce rigorous evidence on whether educational interventions “work.” However, principals and superintendents need evidence that applies to *their* students and schools. This paper examines whether the average impact of an intervention in a particular site—school or district—can be accurately predicted using evidence from a multi-site RCT.

### What Methods Did the Study Use to Predict Impacts?

This paper used three methods to predict the average impact in individual sites: (1) the average of the impact estimates in the other sites, (2) lasso regression, and (3) Bayesian Additive Regression Trees (BART). Lasso and BART used a variety of moderators as predictors, including characteristics of participating students, participating schools, the intervention as implemented, and the counterfactual condition.

### How Was the Accuracy of These Predictions Gauged?

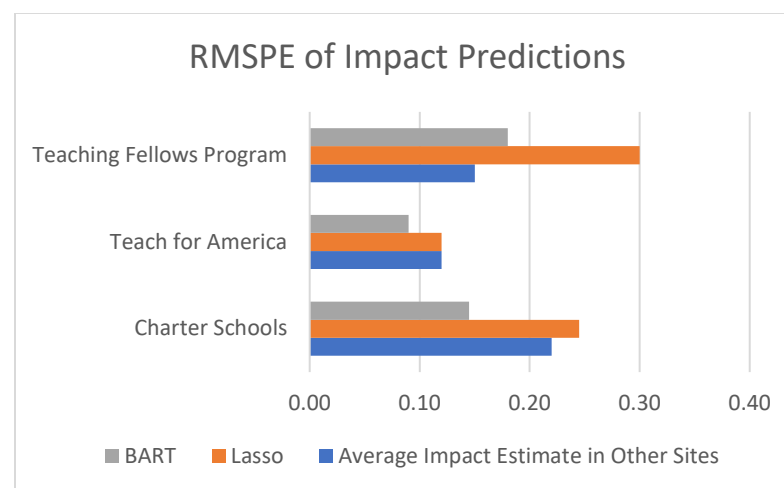
The Root Mean Squared Prediction Error (RMSPE) was calculated for each RCT, prediction method, and outcome variable. The prediction error for each site was estimated by taking the difference between the predicted impact and its impact estimate from the RCT. The RMSPE was calculated by meta-analyzing the estimated prediction errors across all sites.

### What Did the Study Find?

Impact predictions for individual sites were accurate in the two RCTs with little or no cross-site variation in measured impacts (studies of gang violence prevention and student mentoring).

However, in the RCTs with substantial cross-site variation in measured impacts—including studies of charter schools and teacher preparation programs—the impact predictions were not accurate. When data on all four types of moderators were used

to make predictions, BART usually produced the smallest prediction errors (see figure below). However, the RMSPEs were greater than 0.04—the average impact reported in a recent synthesis of evidence on the effects of educational interventions on broad tests of student achievement. This suggests that the prediction methods may either underestimate the true impact in a site by enough to make interventions of average effectiveness seem ineffective, or overestimate the true impact in a site by enough to make interventions with zero impact appear to have average effectiveness.



Note: For charter schools, RMSPE is averaged across four outcome variables.

### What are the Study’s Implications?

The findings suggest that unless the impact of an intervention is similar across sites, multi-site RCTs are unlikely to produce evidence that is accurate enough to help school principals or superintendents predict whether the intervention would be effective in their school or schools.

Full Article Citation:

Robert B. Olsen, Larry L. Orr, Stephen H. Bell, Elizabeth Petraglia, Elena Badillo-Goicoechea, Atsushi Miyaoka, Elizabeth A. Stuart

(2023). Using a Multi-Site RCT to Predict Impacts for a Single Site: Do Better Data and Methods Yield More Accurate Predictions?. *Journal of Research on Educational Effectiveness*:

<https://doi.org/10.1080/19345747.2023.2180464>