Because differences in program features are normally introduced by different operating units, these units become key elements of the design. In fact, the design does not literally compare differences in the program; rather, it compares different sites that happen to implement the program differently. Thus, the design is highly vulnerable to confounding the effect of program variations with the effect of other factors that differ among sites, which may range from regional economic characteristics to the abilities of site staff. The only way to reduce this vulnerability is to have multiple sites representing each programmatic variant.

Implementing this design requires first defining meaningful variants of the program component and then identifying a number of sites that implement each variant. Outcomes are then measured for all or a sample of participants in each group of sites. Participant outcomes are modeled as a function of the program variant they face, their site, and an array of participant characteristics.

Planned Variation Studies

To the extent that program legislation and regulations allow program operators discretion in shaping program components, they also open the possibility for planned variation. In a planned variation design, the agency sponsoring the evaluation (or sometimes the evaluator) arranges for the use of specified variants of the program component by particular sites or in particular circumstances.

If planned variation is feasible, a randomized experiment is likely to be possible and is the preferred design. Individuals or aggregates of individuals are randomly assigned among the variants being tested. Differences in outcomes can be attributed to the differences in the program component. If randomized experimentation is precluded, the possibilities include the same array of designs described earlier for evaluating demonstration modifications to ongoing programs.

Parting Words

This report has noted, at several points, that randomized experimentation is the preferred design for impact evaluation in practically all situations. However, the bulk of the discussion has been devoted to the many quasi-experimental designs that are often used in place of randomized experimentation.

Lest the word count distort the message, we must reemphasize here the importance of exerting all possible efforts to use randomized experiments. For programs that deliver services and benefits directly to individuals and families, randomized experimentation is the only design that, properly applied, is guaranteed to produce unbiased estimates of program impact. All other designs are vulnerable to some bias. Their sources of bias can sometimes be described, but the direction and magnitude of the bias cannot be measured reliably. Thus, all the nonexperimental designs have some substantial probability of producing answers that are far from the truth—which can lead to inappropriate policy decisions that may affect millions of people and billions of dollars of public expenditure.