

WHY RANDOMIZE?



IMPACT EVALUATION

Policymakers, funders, implementing agencies, and consumers all need to decide which programs to support. How can they know which programs work, and which programs work best?

The purpose of impact evaluation is to estimate the effectiveness of a program by comparing what happened to program participants to *what would have happened to these same participants without the program*. Measuring outcomes for program participants is often possible. However, outcomes for the counterfactual—what would have happened to these participants in the absence of the program—can never be observed and thus, must be inferred.

CONSTRUCTING A COMPARISON GROUP

The counterfactual is just an idea—what would have happened to participants in an alternate world in which the program never existed. Because the counterfactual is impossible to measure directly, methods of impact evaluation try to mimic the counterfactual by selecting a comparison group of non-participants who closely resemble the participants. All else equal, the more closely the counterfactual group mirrors the participants before the start of the program, the more confident we can be that any observed differences in outcomes after the program are due to the program itself.

There are many ways to construct a comparison group. For example, the comparison group could be: (a) participants themselves before the program started, (b) those who were eligible for the program but did not participate, (c) non-participants who are either similar to participants, or are dissimilar along observable characteristics that are accounted for using statistical techniques.

The downside with all of these approaches is that they require strong assumptions in order to conclude that any differences in outcomes between the participants and comparison group after the program can be attributed directly to the program. For example, these non-experimental approaches to constructing a comparison group require us to assume that the two groups were equivalent, on average, before the program started (including in ways that cannot be measured, such as their inherent personality traits). Moreover, we would also need to assume that no other factors besides the program affected either groups' outcomes over time. Because it is not possible to test these assumptions, we cannot know for sure whether the changes we observe are caused by the program or by something else.

RANDOMIZATION

Randomized evaluations do not require us to make the assumptions described above. In a randomized evaluation, participants and non-participants are randomly selected from a sample of eligible program participants. Random assignment ensures that, with a large enough sample, the two groups are similar on average before the start of the program. When properly designed and implemented, randomized evaluations produce impact estimates that offer us confidence that any differences in outcomes between the two groups are a result of the program. The ability to isolate the impact of the program from other confounding factors is why randomized evaluations are widely recognized as a particularly credible method for estimating program impact.

Additionally, because randomized evaluations do not require a lot of assumptions, they produce results that are transparent and easy to explain to different audiences.