Why Conduct A Rigorous Evaluation?

The more rigorous an evaluation is:

NEW The more likely it is to tell you something you didn't already know about the program and the more likely it is to produce findings that give you confidence in making decisions about the program (e.g., how to improve it).

The broader the group of stakeholders who will find the results convincing and the more attention the program will get.

The more confidence funders can have in making a funding decision—and the less uncertainty and risk that is involved for funders about whether the program represents a sound investment.

A rigorous evaluation means:

1 Basing your evaluation on a clear logic model that describes **how** the program works to create positive outcomes for participants, systems, or communities.

2 Designing and conducting process and outcome studies that examine whether the program is being implemented as designed and what kinds of impact it has on participants, systems, or communities.

A rigorous *process* study includes:

- Documenting who the program served and how it compares to the intended target population, in terms of characteristics and size.
- Documenting whether each program activity was implemented as planned.
- Documenting any changes or adaptations to the original program plan.
- Assessing how much of each program activity/component participants received (e.g., how many case management sessions were received).
- Tracking program completion and drop-out rates.
- Capturing the perspectives of staff and partners on the implementation process.
- Eliciting participant feedback on why they chose to participate and how well the program met their needs and expectations.

A rigorous *outcome* study includes:

- Sound measurement of outcomes that the program is designed to impact. This should include:
    - **Proximal** outcomes or short-term outcomes that change along the way to produce the end goal (e.g., employment)
    - **Distal** outcomes or the end goal (e.g., re-arrest rates, recidivism)
- Comparison to a "counter factual". This means comparing outcomes for program participants to what would have happened if participants had not received the program. This requires:
    - Statistically testing for differences in outcomes between the program participants and a comparison group.
    - Having the comparison group be as similar as possible to program participants. Ideally, random assignment can be used to generate the comparison group, but when not possible, a waitlist comparison or other nonequivalent comparison group design can work.

For more information, see our video on why rigorous evaluation matters and our video and accompanying infographic on alternatives to random assignment."

Common problems with non-rigorous evaluations:

With a non-rigorous evaluation, you might see differences in outcomes, but these approaches do not include a strong counterfactual. Therefore, you will have no basis for concluding whether those differences can be attributed to the program. *For example:*

Historical comparison

With a historical comparison, such as comparing rates of recidivism at the target facility before and after program implementation, something else might have changed during the period you are measuring that caused the difference in the outcome (for example, a change in local arrest practices or policies).
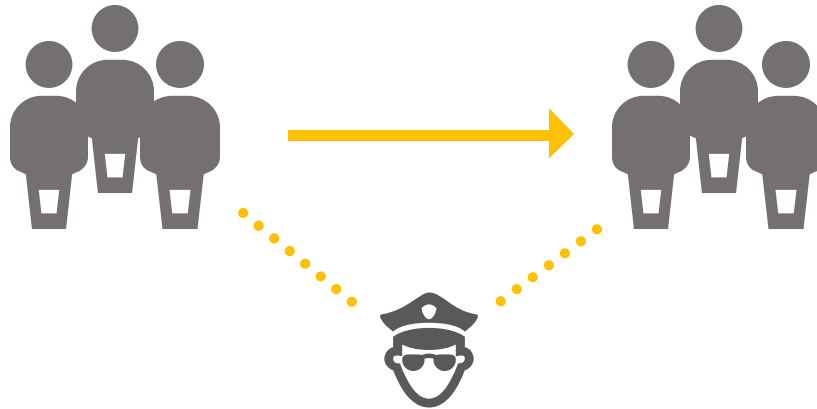
Compare people who opted out

Comparing eligible people who opted to participate in the program with eligible people who chose not to participate or dropped out of the program without completing it is problematic because there might be something about those individuals that both influenced their disinterest in the program and also caused them to have different outcomes.

RTI INTERNATIONAL

Center for Court Innovation