What is program evaluation?

A beginners guide

What is evaluation

This is the sixth in a set of handouts on the key program evaluation questions. This is about whether programs <u>caused</u> outcomes.

Did the program have an effect?

How do you know whether the program improves people's lives?

One commonly used way to find out is to ask whether the program **caused an outcome.** If the <u>program caused</u> the outcome, then one could argue that the <u>program improved</u> people's lives.

On the other hand, if the program **did not cause** the outcome, then one would argue that, since the <u>program did not cause</u> the outcome then the <u>program did not improve</u> people's lives.

How to figure this out?

Determining whether a program caused the outcome is one of the most difficult problems in evaluation, and not everyone agrees on how to do it. Some say that randomized experiments are the best way to establish causality. Others advocate in-depth case studies as best. The approach you take depends on how the evaluation will be used, who it is for, what the evaluation users will accept as credible evidence of causality, what resources are available for the evaluation, and how important it is to establish causality with considerable confidence. (This paragraph suggested by Michael Quinn Patton.)

This handout introduces some of the methods and issues involved in addressing whether a program causes an outcome.

What is evaluation

These are three approaches frequently used to establishing whether a program causes an outcome.

- comparison groups – comparing people in the program to people not in the program

- multiple evaluation methods – comparing results from several evaluations, each using different methods

- in depth case studies of programs and outcomes - showing that the links between what program participants experience and the outcomes attained are reasonable, empirically validated, and based on multiple sources and data. The linkages between program and outcomes are direct and observable. No alternative possible causes offer a better explanation.

The particular method that is used should reflect a careful discussion and understanding of the pros and cons of each method, and agreement among all parties involved.

Comparisons groups and cause:

Comparison groups and random assignment

The idea is this:

Randomly assign people to either be in the program (the 'treatment' group) or to be not in the program (the 'comparison' group).

Since people in the 'treatment group' were randomly assigned, then before the program the two groups of people should be pretty much the same.

Measure the treatment group after they have been on the program and compare them to people in the comparison group.

After the program, if the 'treatment' group people are better off than are the comparison group people, then the difference should be from being in the program, and so it is reasonable to argue that the program caused that outcome.

Additional Resources:

Why do social experiments? Chapter 7 in The Magenta Book, from Policy Hub, National School of Government.
http://www.nationalschool.gov.uk/policyhub/evaluating_policy/

What is evaluation

Advantages and disadvantages of random assignment to treatment and comparison groups.

Advantages:

- The 'treatment' or program effects can be isolated, and other factors can be excluded from causal explanation.

- Provides results that are easy to explain.

Disadvantages:

- Often not practical to do. Can't randomly assign people to program or not program, and may be unethical to randomly assign someone to no treatment.

- Randomly assigning people to be in the program is not how programs really work, so results of the <u>evaluation</u> may not apply to the program as it <u>really exists</u>.

- Can't be applied to causes that operate over the long term or to programs that are very complex.

- Can tell whether a program caused outcome, but doesn't give much in depth information about **why** or **how**.

- People in treatment group know they are getting treatment so outcome may be due to knowledge, not to treatment.

This is a summary of points from:
(Munck and Jay Verkuilen) "Research Designs,"
http://www-rcf.usc.edu/~munck/research/methods.html

The last point is one of many from:
A Summative Evaluation of RCT Methodology: & An Alternative Approach to Causal Research. Michael Scriven. Journal of MultiDisciplinary Evaluation, Volume 5, Number 9. http://jmde.com/

What is evaluation

<u>Comparison groups and non-random assignment</u> When random assignment is not possible, <u>quasi-experimental design</u> can be used. In this method, people "are not randomly assigned to groups but statistical controls are used instead."

Quasi-experimental designs. In Statnotes: Topics in Multivariate Analysis, by G. David Garson   http://www2.chass.ncsu.edu/garson/pa765/design.htm#quasi

There are several versions of this approach:

- Comparing people already on the program to those who are not on the program. One example is to observe (O) people before they join the program or there is an intervention (X), then observe both groups after :

Pretest-posttest design
-Intervention group O_{before} X O_{after}
-Comparison group O_{before} O_{after}

- Measuring the client many times before they join the program (or before a new intervention) and many times afterward, them compare before to after. One example is:

Time series design
-Intervention group O_1 O_2 X O_3 O_4

- Combination of the two above

Time series design
-Intervention group O_1 O_2 X O_3 O_4
-Control group O_1 O_2 O_3 O_4

What is evaluation

A <u>major challenge</u> to non random assignment approaches is that people <u>on</u> the program may start off being very different from the people <u>not on</u> the program.

That is, only <u>some</u> people choose to be on the program. Something made these people different and it may be the something which caused the better outcome, not the program.

One way to deal with this is to collect as much information as possible on characteristics of the people and program that relate to the program outcome, and use this information in statistical analysis to "control" for the differences between people on the program vs people not on the program.

The problem is that there may be differences, some critical, that are not observed, and for which the evaluator has no data.

Additional Resources:

AllPsych On Line. By Dr. Christopher L. Heffner
Section 5.3 Quasi-Experimental Design
http://allpsych.com/researchmethods/quasiexperimentaldesign.html

Quasi-experimental designs. In Statnotes: Topics in Multivariate Analysis, by G. David Garson
http://www2.chass.ncsu.edu/garson/pa765/design.htm#quasi

<u>Design diagrams on previous page from</u>:
Measuring the Difference: Guide to Planning and Evaluating Health Information Outreach. Stage 4, Planning Evaluation. National Network of Libraries of Medicine
http://nnlm.gov/evaluation/guide/

What is evaluation

Multiple evaluation methods could support the idea that the program causes the outcome if different sources agree.

For example, collect information from:

- Program participants
- Program staff
- Community members
- Subject experts
- Published research and reports

Collect data through many methods, for example:

- Surveys
- Interviews
- Observations
- Program administrative data

If data from different sources don't agree, it doesn't <u>necessarily</u> mean the results from any of the sources are not valid. However, the more agreement there is from different sources, the more confident you can be about your conclusions.

Additional Resources:

An Introduction to Mixed Method Research. By Jennifer Byrne and Áine Humble. Atlantic Research Centre for Family-Work Issues. Mount Saint Vincent University.
http://www.msvu.ca/site/media/msvu/MixedMethodologyHandout.pdf

An in depth case study can be used to demonstrate the connection between the intervention and the outcome.

An in-depth case study documents in detail what a group of participants experienced in a program and any ways in which they have changed so that the evaluator and users of the evaluation can make a judgment about the likelihood that the program led to the observed changes. For example, a group of chronic alcoholics go through a residential chemical dependency program. Their participation is fully documented. They return home maintaining their sobriety. They attribute their sobriety to the program as do their families, friends, and program staff. These multiple sources agree on the documented causal chain. The links between what they experienced and the outcomes attained are reasonable, empirically validated, and based on multiple sources and data. The linkages between program and outcomes are direct and observable. No alternative possible causes offer a better explanation. The preponderance of evidence points to a conclusion about causality. Establishing causality involves both data and reasoning about the findings. (This paragraph contributed by Michael Quinn Patton.)

Conclusion:

Random assignment is often seen as a very clear way to show whether a program causes an outcome. However, random assignment is often not practical or reasonable.

Other methods such as non random assignment, multiple evaluation methods, or in depth case studies are more practical and can be used to give reasonable arguments about whether a program caused an outcome.

However, these methods are less certain in establishing that the program is the cause of the outcome. There may be other things going on that are unknown and these other things might really be the cause of the outcome. It is more difficult for these other methods to rule out other possible causes, although the other methods can, again, establish reasonable likelihood.

If a cause cannot be established, the evaluation can be used to describe what happened.

For example, the evaluation could say, "After the program, people were better off." This doesn't necessarily mean it was the program that made the people better off, but the program may be one reasonable cause. Gathering more evidence, from multiple methods, driven by a very clear understanding of the program, can help determine the most reasonable explanation of causes of the outcomes.

What is evaluation

Additional Resources:

Steps in Program Evaluation
Gather Credible Evidence
http://www.cdc.gov/eval/steps.htm#evidence

Program Evaluation: A Variety of Rigorous Methods Can Help
Identify Effective Interventions
GAO-10-30 November 23, 2009
http://www.gao.gov/products/GAO-10-30

What Counts as Credible Evidence in Applied Research and
Evaluation Practice? Stewart I. Donaldson, Christina A. Christie and
Melvin M. Mark.   http://sites.google.com/site/credibleevidence/
Especially see the free resources, Donaldson, S.I. (2008). A
practitioner's guide for gathering credible evidence.

What Constitutes Credible Evidence in Evaluation and Applied
Research? Claremont Graduate University Stauffer Symposium
http://www.cgu.edu/pages/4085.asp

A Summative Evaluation of RCT Methodology: & An Alternative
Approach to Causal Research. Michael Scriven. Journal of
MultiDisciplinary Evaluation, Volume 5, Number 9.
http://jmde.com/

**The most recent version of this handout is always at
http://gsociology.icaap.org/methods/**