HEALTH PROGRAM EVALUATION – DESIGNING THE STUDY

Dr Joseph Antonio Molina
Principal Research Analyst
Health Services & Outcomes Research
National Healthcare Group

What is an Evaluation Design?

Design refers to a plan for meeting an objective. While a research design is a blueprint for conducting a study, an evaluation design is the detailed strategy for conducting an assessment of a health program. The designs used in program evaluation are based on epidemiological methods.

What are the requirements of an evaluation design?

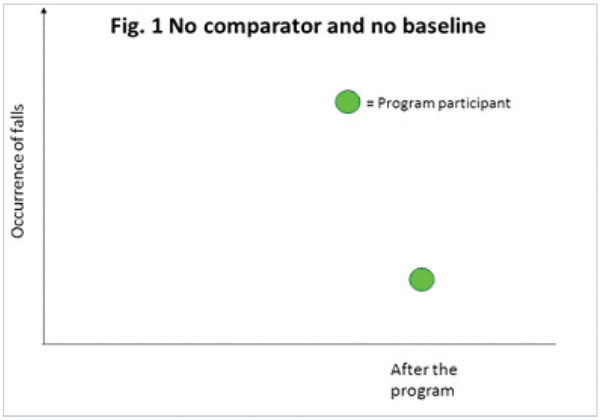
1. Complete and accurate data – While research studies need complete and accurate data, this may be difficult to achieve in real-world research. This is the motivation for using prospective evaluation designs where data are longitudinally collected from program start (baseline) to program completion (follow-up). It is easier to ensure completeness and accuracy while data is being collected than after the data has already been gathered.

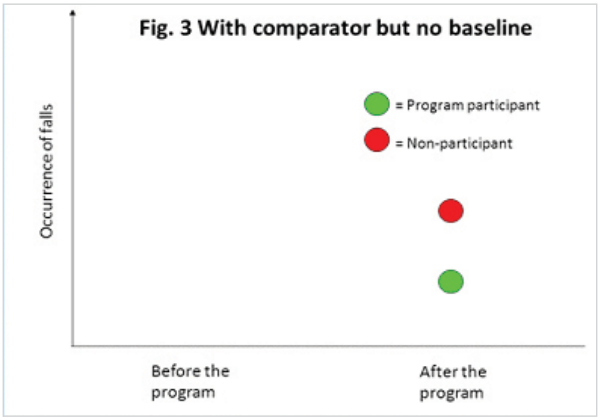
2. Basis for proving that the program is "effective" – As program evaluation aims to measure how well a program is performing, the best way to establish this is by comparing outcomes between:

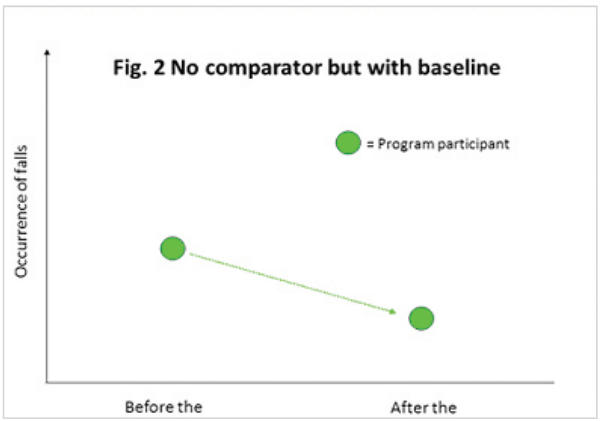
    a. A group of program participants versus a group(s) who are not; and
    b. Baseline and follow-up measurement for both groups

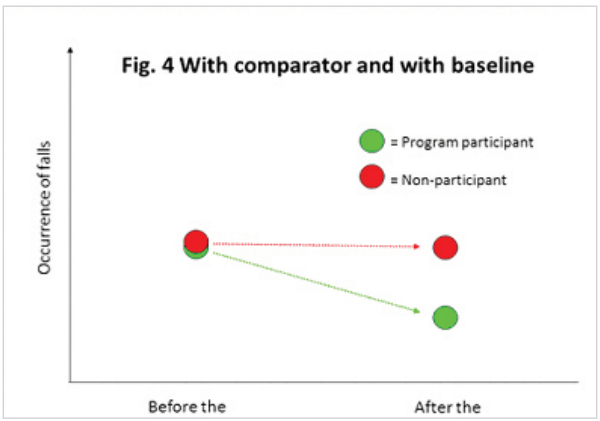
Figures 1 to 4 show hypothetical "evaluation" results from a falls prevention program. When there is no comparator and baseline data, Figure 1 provides the least information to make an assessment of the program outcome.

Although Figures 2 and 3 have additional data with which to compare the occurrence of falls during follow-up among program participants, it is still not enough to make a complete assessment. Figure 4 provides the most complete data for assessing program performance.









On the issue of comparability of treatment groups: Similarity in baseline characteristics between participants and non-participants facilitates attribution of program effects. However, there are post-design analytical approaches which can adjust for baseline differences. These include propensity score matching, use of instrumental variables, difference-in-difference analysis and regression-discontinuity design.

Common evaluation designs:

1. Before-after, program-no program design

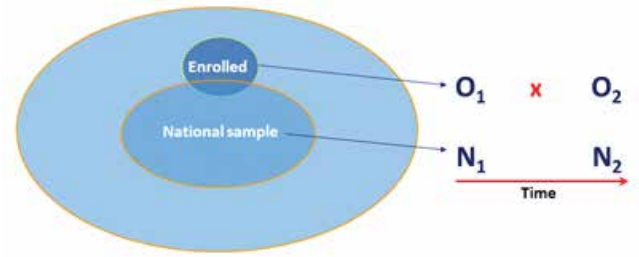
Comparisons are made within (before-after) as well as between groups (program participants versus non-participants), hence it is possible to attribute effects to the program.

2. Chronological series (single group time series)

Multiple serial assessments are done before and after program implementation; trends in the outcome before and after program implementation are compared. Due to the absence of a non- program group, there is no way to discount the effects of adaptation to repeated assessments (testing effect), exposure to interventions from sources other than the program itself, or natural progression of the condition (maturation effect).

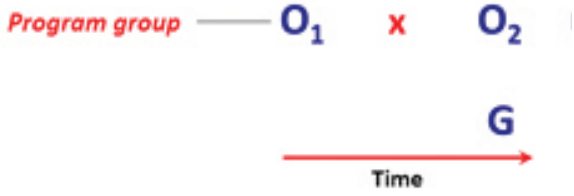
3. Before and after studies using national averages as comparison

Similar to design 1, but comparison is with a national sample which may include some program participants.



4. Goal-based evaluation model

Targets for the outcome are determined before program implementation. The program is held accountable to prior expectations rather than to relative performance against an actual comparison group. Setting targets for program performance require a strong logic model.



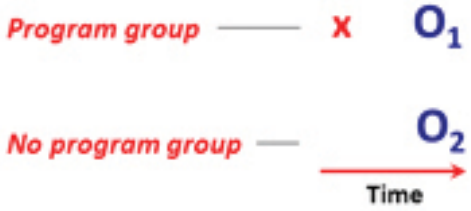
Note: G = Pre-determined program goal/target

5. Single group before-after (pre- post-) "design"

No parallel control group; hence results are more suggestive rather than conclusive of program performance. Aside from the limitations of the chronological series design, the single pre-post- design is susceptible to regression to the mean (which implies that at their extreme states, some diseases/conditions may regress to less severe states even in the absence of any intervention).

6. Post-test only "design" with non-equivalent groups

No baseline data, hence it is not possible to rule out maturation effects. There is no way to assess comparability of groups at baseline.



7. One-group post-test only "design"

The weakest of all "designs," this evaluation is entirely uninformative except to describe the state of participants after enrollment into the program.

Of the above mentioned designs, the last three should be avoided as they are susceptible to many biases. Users of potentially misleading results run the risk of making flawed conclusions about the program.

Evaluation designs can be hybrids of established designs. Ultimately, the program team should aim for a design which can generate accurate and relevant information to aid in decision-making.