Developing a Sampling Plan

Preface

There are some key threads that provide structure to the intertwined decisions in an Evaluation Plan. These are:

- **What do you want to know about your program?** This is a key determinant of the Evaluation Questions, and is itself shaped by the program's structure and logic, by stakeholder interests, and by program development needs.

- **Closely related: What kinds of claims do you want to be able to make?** [1] These considerations help sharpen the Evaluation Questions, and influence the choice of Sample (who will you get information from?), Design (what will be compared to what?), Measurement strategy (what will you be looking for?), etc.

Key questions in determining your Measurement Strategy:

- **What would provide "evidence" of what you are trying to assess?**

- **How might you gather that evidence?**

So it's all about gathering evidence that will be useful for YOUR purposes. These purposes might be purely internal, if you are focusing on program development and improvement; or external if you need to communicate to stakeholders about program activity and/or effectiveness; or both. **"Sampling",** or selecting the group from which to gather information, extends this decision-making process, to make sure that the evidence you gather will be able to support the kinds of claims you want to be able to make, or conclusions you want to be able to draw.[2]

Key terms in Sampling:

A. "POPULATION" (or "population of interest"): *The "population" is the entire group that you are interested in learning about or making claims about.*

1 Obviously, regardless of the "claims you might want to make," the evaluation might give the opposite answers. This is still valuable information, as it will promote program improvements or a better allocation of scarce time and resources, etc.

[2] There are a number of good references on sampling that supplement what is discussed here. Taylor-Powell, E. (1998) *Sampling,* Cooperative Extension Publications, University of Wisconsin, Madison, WI. (http://learningstore.uwex.edu/Sampling-P1029C237.aspx) covers most of the concepts here and offers careful explanations and how-to instructions for obtaining different types of samples (random samples, purposeful samples, etc.) Another excellent reference is http://www.socialresearchmethods.net/kb/sampling.php .

*If the data are being gathered from non-participants or from objects (databases, or physical records for example), then the "population" is the entire set of sources related to or produced by the group of interest.**

***A subtle distinction to clarify:**

In a lot of evaluation writings and discussion, the terms "population" and "sample" are described only in terms of sets of *people* rather than (as included in the definitions above) sets of data. If your measurement strategy involves getting data about people from the people themselves (participants, for example) then it is appropriate to describe your sample as consisting of the subset of people.

However, it is important to keep in mind that your sources of data might be different from the people you are actually interested in. For this reason the definitions above include the language about sets of data. For example, it might be a good measurement strategy to obtain data about the people you are interested in by obtaining it from *other people* (observers, teachers, etc.) or from *objects* (such as journals, record books, food prepared in a cooking demonstration, etc.)

> **Illustration:** Suppose you are interested in evaluating the post-program equipment handling skills of your participants. Suppose your measurement strategy is to test the pieces they produce from using the equipment (in order to assess consistency, quality, amount of material wasted, or whatever is of interest). The *population* in this example would consist of all the products produced by the participants in the program you are studying. The *sample* would consist of the selected subset of the products you actually intend to test.

Note: This distinction is an example of how the decision about measurement strategy interacts with the definition of sample and the decision about sampling.

Examples (these will be carried forward in subsequent sections):

(1) For an early lifecycle evaluation focusing on program implementation (such as participant reactions to the program that was offered) or program outcome (such as post-only knowledge of the participants in the spring of 2010), the "population" could be the full set of participants in the program in the spring of 2010.

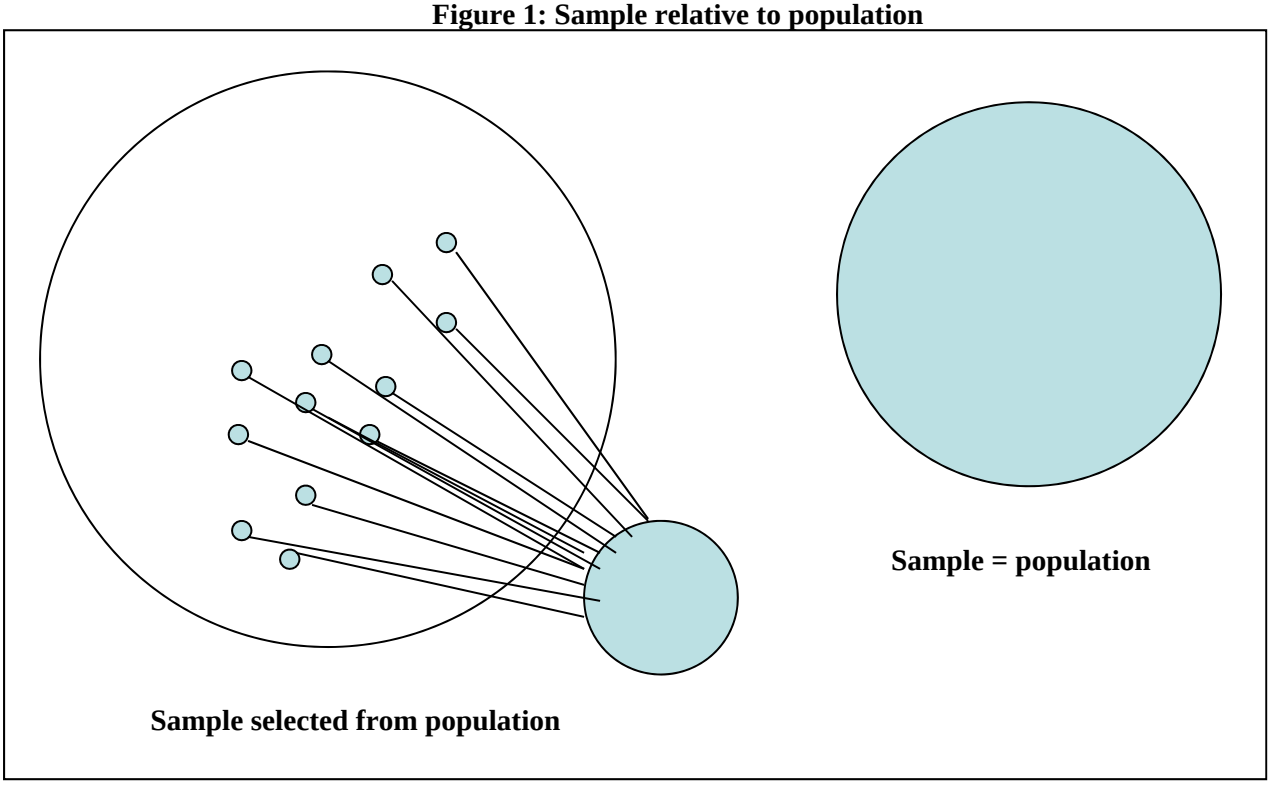
(2) For a later lifecycle evaluation that is intended to yield conclusions about the program's likely benefits for future teen participants in the community, for example, the population could be all eligible teens in the community.

(3) For an evaluation that is intended to assess which program elements were most effective for the high achievers in past sessions of the program, the population could be all the high achievers in all past (presumably you'd have a date range of past sessions you wanted, and you'd have a definition of what you mean by "high achievers").

B. "SAMPLE": *The "sample" is the set of individuals you intend to get data from, or the set of data you intend to collect in order to conduct the evaluation.* The sample is either equal to, or is a portion of,

or is systematically related to, the population of interest. The process of figuring out what the sample will be, relative to the population, is the sampling decision.

Figure 1: Sample relative to population



Examples (cont'd from above):

(1) For this evaluation, the sample could be the whole set of participants (sample = population). However, if the group is large, or if time is just too limited to include everyone, the sample could be a selected subset of the participants. How you select that subset, and how it relates to the overall population, is what has to be decided in sampling.

(2) For this evaluation, where there is a need to "generalize" the evaluation results in order to make claims about the program's likely effect on individuals who have not yet participated in the program, the sample might be all or some of the recent program participants, selected so that they will be "representative" of all eligible teens in the community. So, there has to be a systematic relationship or correspondence between the characteristics of the sample and the characteristics of the population.[3]

[3] In order to draw really strong conclusions you'd have to address the fact that, by definition, the program participants are "different" from other teens in the County because for reasons you might or might not know, these youth chose to do the program and the others did not. So you might need to control for that unavoidable bias by incorporating a matched comparison group of non-participants. But it will often be sufficient for your purposes to make less aggressive claims and be able to say, for example, that future participants are "likely" to experience the same benefits.

(3) For this evaluation, where you are not necessarily attempting to generalize, you still need to do "purposeful" sampling. This example would involve "outlier" or "extreme case" sampling – deliberately selecting the "high achievers" from among past participants, and selecting the sample according to a systematic process. If the measurement strategy is to do intensive case studies then your sample might be quite small.[4]

Note: The term "sample" refers to the set of sources from which you intend to get information. Because of limited response, the actual sample you end up with may be smaller than your intended sample. This potential shortfall should be taken into account when you decide how large your (intended) sample should be, so that in the end you can expect to have "enough" data to make the assessments you wish to make. There are also things you can do to try to increase the "response rate" so that you don't have too many incompletes or non-respondents.[5]

The Sampling Decision – what to consider:

There are options, and choices to be made when sampling. As noted in the examples, the sampling decision governs how the Sample relates to the Population, and from there, what claims can legitimately be made about your evaluation results. It also governs how you will go about selecting the actual sample.

If you wish to *generalize* your results, then you will have to ensure that your sample is representative of the larger group to which you wish to generalize. Depending on how strong a generalization claim you want to make, you will have to engage in varying degrees of rigor in ensuring "representativeness", or the extent to which conclusions drawn about the sample can be said to represent or speak for the population. It may be sufficient to ensure that your sample roughly matches the population on a simple list of characteristics (demographics, etc.) in order to claim a "likely" result or to claim that the fact that your program is "associated" with certain outcomes suggests that these benefits would occur for others who are similar.

However if you want to make a stronger claim that your program caused the effects you observe and to begin to prove that it would be likely to cause them in other members of the population of interest, then you would need to obtain more detailed information about your sample and select them randomly from the population of interest. You would also need to select a comparison group from the population of interest. To make a full claim of causality and generalizability you would need to randomly select participants and assign some to the program and some to a "control" group. Situations like this are beyond the scope of this document.

Sample size is a particular element in the sampling decision that matters if you want to draw conclusions you want claim also hold for the larger population of interest. In general, to assure that the sample fairly represents the population, the sample should be above a certain percentage of the population. How large

[4] For a list of purposeful sampling strategies, see Table 1 in Taylor-Powell, p. 7, which adapts a table from Patton, Michael Q. *Qualitative Evaluation and Research Methods* second ed., Newbury Park, CA; Sage Publications.
[5] For more on response rates and how to adjust sample sizes accordingly see Ellen Taylor Powell's "Sampling" referred to earlier. Also see "Random Samples: How and Why?" in the Pennsylvania State Extension's Evaluation Tip Sheets, at http://extension.psu.edu/evaluation/pdf/TS57.pdf .

that percentage needs to be depends on the size of the population. Generally, the smaller the population, the larger the sample percentage needs to be, because smaller populations tend to have more variability. For guidance on deciding sample size, see Appendix 2 in Taylor-Powell, "Sampling", which includes a table of recommended sample sizes for different precision levels.

If you do not need to generalize your results, but are instead interested in more in-depth information about a program, about participants or about a delivery method, you may want to do *purposeful sampling*, in which you select cases likely to shed light on a question or issue of interest. In this type of sampling, the above considerations on sample size do not apply. It will simply be important to explain and justify the sampling decisions you make.

For both kinds of sampling, random and purposeful, it is important to consider *feasibility*. If you find you have chosen an approach or sample size that is not feasible given the time frame and/or staff resources available, it may be necessary to scale back. So it is best to think this through **before** finalizing sampling plans.

How-to: See Taylor-Powell for how-to steps on basic sample selection methods including simple random sampling, systematic samples, etc. See also the Penn State Extension and Outreach "Evaluation" web page, which also has a number of resources on sampling, including practical, step-by-step directions, at: http://extension.psu.edu/evaluation/sampling.html.

What to put in the Sample Section of the Evaluation Plan

In the Sample section of your Evaluation Plan for each evaluation question, briefly describe the population of interest, and say whether you intend to include in the sample the entire population of interest, or generalize from a smaller sample to a larger population, or do purposive sampling. Give a rough idea of your thinking about sample size. If selecting a subset from the population, indicate how it will be selected. (If you only have time to evaluate half of the products, will you select every other one? Or draw numbers from a hat in order to see whose products get tested and whose don't? Will half be enough to give you a reliable answer? And so on.) If you will use a comparison group, give the above kinds of information for that group as well.