appendix D

Sampling and Sample Size Guide / Logistics Guides

Types of Sampling

There are several different methods that can be used to select a sample of people or things to be included in an evaluation. The bottom line is that you want the sample to represent the whole group so that you can apply your findings to the program. It is wise to discuss your particular situation with a program evaluation specialist or epidemiologist. There are several methods that can be used in the public health setting. They all have advantages and disadvantages. Select the one most appropriate for your evaluation. The five following methods are frequently used in public health evaluations.

Simple Random Sampling

A list of all the people/things is made and then individuals are selected randomly to be included in the sample. The main advantages are that everyone in the whole group is given an equal chance of being included in the study; statistical tests are mostly designed for this type of sample. The disadvantage is that you need a complete list of everyone in the group, which is often unavailable. If there is great variation among the group then, your random selection might give you a sample that inaccurately represents the whole group. If a list is not available, then you can modify this approach and randomly select people or things as they become known to the program.

Stratified Random Sampling

This method is similar to random sampling, but there is an additional step. All members of the whole group are divided into "strata" or "groups" (e.g., men and women). Then participants within each strata are randomly selected to be included in the sample (e.g., 10% of the men and 10% of the women). The advantage is that there is a greater chance the sample will reflect the underlying distribution of the entire group's characteristics. The choice of strata depends on what is being measured. In general, you should stratify by characteristics that will influence the findings. For example, if women smoke less than men, then you should stratify by gender when measuring smoking rates in the population. The main disadvantage of this method is that, as with simple random sampling, you need a complete list of everyone in the whole group before you begin, and this is often unavailable.

Systematic Sampling

If a complete list of people/things is not available, then you may decide to select every second or every fourth person/thing to be part of the sample. This is systematic sampling. The advantage is that it is easy to do. The main disadvantage is that there may be a bias in how the people/things are selected if there is a bias in how the people/things are ordered. For example, if you select every second person who enrols in the Parenting Program to participate in the sample, and all the women register first and their husbands second, then you will only get women in the sample.

Cluster Sampling

There are natural groupings within the population — examples are, schools or neighbourhoods. In cluster sampling, the group as a whole is divided into these natural clusters and then specific clusters are selected, ideally randomly, to be included in the sample. Everyone within the selected clusters is included in the sample. The evaluation is much easier to do if you can take advantage of these natural groupings. This works well if there is homogeneity among the clusters. If the clusters are quite different, then it is wise to stratify the clusters first and then randomly select from within the strata. The disadvantage with this method is that the chosen clusters may not reflect the true underlying population.

Convenience Sampling

In this sampling method, you include volunteers or people/things to whom you have ready access. This can be useful when conducting pre-tests, but it is not wise for your program evaluation itself. There is no way to assess if the people/things included in the sample in fact represent the underlying population.

Sample Sizes

The findings from a sample of the whole group provide an *estimate* of the findings for the group as a whole. Different samples will produce different findings. For example, if you took three samples of 100 participants in the Parenting Program to measure their satisfaction with the program, the findings might be as follows: group 1 - 76% very satisfied; group 2 - 81% very satisfied; group 3 - 79% very satisfied. The findings are similar, but not exactly the same. This is not surprising because there is variation in how the participants will rate the program. The findings for the sample depend on the mix of individuals within it. This is why it is critical that you use a sampling method that ensures your sample is as similar to the whole group as possible.

Given that samples only give an estimate of the findings of the whole group, it is a good idea to calculate the confidence interval associated with findings from a sample. The confidence interval tells you how confident you can be that the findings from the sample reflect your whole population. For example, the 95% confidence interval for the estimate of the satisfaction for the Parenting Program of 79% would be +/- 8%. This means that you are 95% sure that the satisfaction with the program among the group as a whole is somewhere between 71% and 87%. The confidence interval depends on the size of the population as a whole, the size of the sample, and the underlying distribution of the characteristic in the population. The formulas used to calculate the confidence intervals include these parameters.

It is best to consult an epidemiologist or program evaluation specialist prior to selecting your sample size so he/she can help you decide how many to include in the sample. The sample size is the one parameter you control that will influence the confidence interval. You must feel comfortable that you will get as precise an estimate as you need before you start the evaluation. The need for precision will vary, depending on the situation.

Some examples of confidence intervals for varying sample sizes, if the finding is that 50% were satisfied with the program are:

sample size	**95% confidence interval**
20	+/- 22%
50	+/- 14%
100	+/- 10%
150	+/- 8%
200	+/- 7%
300	+/- 6%
500	+/- 4%