CHAPTER 5

Survey Methodology & Sample Design

CHAPTER 5
Sampling Design Methodology

5.1 Introduction

This section provides a general overview of the sample design methodology recommended for use by the area aging agencies (AAAs) involved in the Performance Outcomes Measures Project (POMP). This section discusses the objectives, the process of designing the sample and selecting the sample.

Designing and selecting a good sample is one of the first steps in conducting a survey. A good sample design also insures that the data collected and results will be statistically significant.

POMP will have somewhat different sample designs and selection procedures for each performance measure. These specific sampling instructions are contained in the implementing instructions accompanying each performance measure questionnaire. This section is provided to give the reader a general understanding of the importance, structure and use of sampling.

5.2 Sample Design Objectives

The primary objective of the sample design is to select a subset of the population being served who will reflect the views of the entire population within a certain percentage of confidence. It is much faster and less expensive to interview a subset of the population than to try to interview the entire population being served. In order for this subset to accurately reflect the views of the entire population, a sample of the population has to be carefully selected so that the survey results are statistically significant.

5.3 Sample Design and Selection

The quality and usefulness of results from a survey depend largely on the procedures used to select a sample of clients. Because surveying every client in an area agency, region, or state is usually impossible, impractical, and unnecessary, a sample of the entire population should be selected. Samples can be selected in many ways. The results from a good sample can be generalized to the entire client population from which the sample was drawn. The results from a poor sample may only refer to the clients who participated.

The two main factors contributing to a good sample are sample size and sample selection. Any estimate that is based on a sample will have some degree of error in it. Sample size is directly related to the magnitude of the error expected in the results. The magnitude of the error will reflect your confidence in the statements that you make about the population based on your sample results. Typically prior to conducting a survey, an agency would decide on the maximum amount of error allowed in their estimate of the population. Deciding on the

margin of error involves marking off the limits within which the population mean probably lies, and determining the probability of selection. In technical terms, you would state the level of confidence of your hypothesis that the population mean resides within the specified confidence interval.

The second issue of sample selection is concerned with the method for selecting the clients to be included in the sample. To be able to generalize findings to the population, a random sample of clients should be selected. A random sample is one where every member of the population has an equal chance of being included in the sample and each selection is made independently of all the others. This provides a representative and random sample. However, a representative sample is not always a random sample.

This section describes a one tier sampling process, which is the simplest to use. It describes how to select an appropriate sample size and how to randomly select members from the population to be included in the sample.

5.4 Sample Size Estimation

There are several different procedures that can be used to produce sample size estimates. Each method requires assumptions to be made about the characteristics of the population. The method used in this study is based on the confidence interval equation and assumes that the results of the data analysis will be a distribution of the percentage of responses by question category. It is not intended as an estimation tool when an average response is to be derived.

The confidence interval method of estimating the sample size requires knowing the actual size of the population of interest, an estimate of the answers expected on the survey (advanced answer estimate), and an estimate of the amount of error that will be tolerated in the final performance measure estimate. The size of your sample will be directly related to these factors.

The size of the population of interest should be known after the master client list is generated. The advanced answer estimate refers to the answers expected on the survey. Since most of the questionnaires being tested are new, there is no previous data on how people will respond on the different questions. Therefore, the worst case assumes that 50% of the people will respond with the same answer on a question. For example, when there are two possible answers (e.g., Yes/No), it is assumed that 50% of the respondents will say Yes and 50% will say No. The importance of the advanced answer estimate is that it can reduce the sample size requirements.

The amount of error refers to the confidence interval about our estimates. The confidence interval is the plus-or-minus figure usually reported in opinion poll results to indicate the range where the population estimate would most likely fall under. For example, if 42% of the sample reply with a Yes to a question and the confidence interval is +/- 5, then we can be highly confident that the true population percentage would be between 37% and 47%. For this study, a confidence interval of +/- 5% is preferred. However, since this is a pilot study

and the period of data collection may not allow grantees enough time to collect the required sample sizes, a confidence interval between +/- 5% (or a 10% spread) and +/- 10% (or a 20% spread) will be allowed.

Another possible source of error that the sample size estimation procedure has to account for is the non-response rate or the percentage of clients who will refuse to participate, cannot be located, or other reasons. To insure that the number of completed questionnaires matches the number needed for the desired level of precision, the initial sample size estimate needs to be adjusted by the percentage of clients expected to refuse or cannot participate for other reasons. To make this adjustment, an estimate of the non-response rates is needed. The response rate usually is estimated from previous surveys conducted by the local agency and will vary by method of administering the questionnaire.

For POMP, a Sample Size Calculator program has been developed to assist agencies in determining their sample size requirements. Table 1 is an example that highlights how sample sizes vary by population size and precision levels. Regardless of population size, it is recommended that the minimum sample size should be at least 50.

5.5 Random Selection Methods

Randomization is the process that ensures that each client in the group of interest has an equal chance of being sampled and surveyed. In order to make reliable inferences about the group of interest (e.g., congregate meal clients) from which the sample is drawn, it is extremely important for the sample to be representative of the population of interest, and random selection insures that this occurs. Since POMP is a demonstration study with a limited data collection period, which does not extend to the whole year, the selection of clients may not be representative of the population over the year. For example, in POMP only callers to the Information and Assistance Services during a two month period, such as June and July of 2000, were sampled, but the findings based on these callers during this two month period will be generalized to the entire population of callers over a 12-month period. Technically, for an annual estimate, sampling and selection of I&A callers should be done across all 12 months to avoid unknown biases in the results, such as differences in the staff personnel throughout the year.

There are two popular methods for randomly selecting a sample of clients. The simple random sampling method, which is the most popular method, selects clients from a list of all clients based on a random number assignment. The systematic sampling method selects clients based on a sampling rate, such as, every fifth person who enrolls in the program will be included in the sample and given the questionnaire.

The method selected will depend on several factors such as how the information on the clients is stored (i.e., paper logbooks versus a computer database) and whether the population of clients is known prior to data collection. In some instances where the population of interest is very small (e.g., newly enrolled nutrition program clients), no random selection method may be needed; instead, all clients in the population of interest would be administered the questionnaire.

Table 1. Sample Size Estimates Based on Population Size and Confidence Interval*

	Population Size										
Confidence Interval	50	75	100	150	200	250	300	350	400	450	500
+/- 1%	50	74	99	148	196	244	292	339	385	432	477
+/- 2%	49	73	96	142	186	228	269	309	347	384	420
+/- 3%	48	71	92	133	171	206	239	270	298	325	350
+/- 4%	47	67	87	122	154	181	206	229	249	267	284
+/- 5%	45	64	81	111	136	157	175	191	205	218	229
+/- 6%	43	60	75	99	119	135	148	160	169	177	185
+/- 7%	41	56	68	89	104	116	125	133	140	146	150
+/- 8%	39	52	62	79	91	99	106	112	117	121	124
+/- 9%	36	48	57	70	79	86	91	95	98	101	103
+/ 10%	34	44	52	62	69	74	78	81	83	85	87

	Population Size										
Confidence Interval	550	600	650	700	750	1,000	1,500	2,000	5,000	10,000	15,000
+/- 1%	523	568	612	656	700	913	1,313	1,681	3,390	5,128	6,186
+/- 2%	455	489	521	553	584	725	956	1,137	1,724	2,083	2,239
+/- 3%	374	397	418	438	457	539	657	738	948	1,047	1,085
+/- 4%	300	314	327	339	351	397	458	495	581	617	630
+/- 5%	239	248	256	263	270	297	329	348	388	404	410
+/- 6%	191	197	202	206	211	226	245	255	276	284	287
+/- 7%	155	158	162	165	167	177	188	194	206	210	212
+/- 8%	127	129	131	133	135	141	148	152	159	162	163
+/- 9%	105	107	108	110	111	115	120	122	127	128	129
+/ 10%	88	90	91	92	92	95	98	100	103	104	105

* Table based on a 95% confidence level and a 95% responses rate.

Random Sampling Method

The simple random sampling method would work best for agencies who want to sample from the Master Client List. The steps for selecting clients to be included in the sample would involve (1) assigning a random number to every client in the list, (2) sorting the list in ascending order by the random number, (3) selecting the first *"N"* records, where *"N"* is the sample size estimate, and (4) copying these records to your log book. These selected records in the sample list will be the clients who are contacted and administered the questionnaire.

An example of using Microsoft Excel to select a sample is given here. We'll assume that we have a population of 600 transportation clients and have to obtain a sample of 90 client names from this population. The first step would be to import your list of clients into Excel, such that each client is assigned to one row in the spreadsheet. Next would be to assign a random number to each client by using the RAND function, i.e., create a new column and set each cell in this column to the value "=RAND()".[1] Next, place your cursor on the first name in the list and select the sort function under the data menu. The sort function will automatically highlight all 600 names in your list, just verify that it contains all the records you want to sort. Under the sort menu, it will ask you which column you want to sort by, select the column with the random number. You can sort in ascending or descending order, it doesn't matter. Once you resort the data, select the first 90 clients in the list for your sample. Since the whole list is now randomly sorted, you could select the first 90 clients on the list or the last 90 or a group of 90 in the middle for your sample.

Systematic Sampling Method

The systematic sampling method or the sampling rate method is useful for sampling historical records that are in paper logbooks or for sampling clients from an unknown population, such as persons calling for Information and Assistance service during the next two months. This sample selection method determines a sampling rate that will be used to select clients from the sample. For example, the sampling rate may indicate that every fifth person calling for information should be included in the sample.

Systematic sampling is fairly easy procedure. The first step involves determining the sampling rate by dividing the number of clients in your population by the required sample size. For example, if the population is 600 transportation clients and if the sample size is 90, then the sampling rate is 600 divided by 90 or 6.7, which should be rounded down to every 6^{th} client. Once the sampling rate is determined, you need to determine a starting point. Starting with the first client on the list is not recommended; instead, randomly select a number between one and the sampling rate number (e.g., a number from 1 to 6). This will be the starting point and the first client selected for the sample. To select the next client add the sampling rate to the previous record number. For example, if 2 is the starting record, then the next client selected would be 2 + 6 or the eight record and the following would be 8 + 6 or

2 Once you assign a random number to every client, you might want to code the value, since the RAND() will change its value every time you make a change or edit to the Excel workbook. To hard code the value, highlight the random number column, copy it to your clipboard, keep the random number column highlighted, and perform a paste special command that will only paste the value of the cells.

the 14th record. This process is continued until the end of the list of clients. The following illustrations highlight this process for selecting names on a list or out of a stack of records.

Transportation Clients

Client ID	Client Name	Phone No
1101	Cathy Crispin	212-8268
✓1103	Charles Irwin	212-4967
1104	Billie Bob	214-5188
1106	Sherlie Temple	214-6299
1109	John Rogers	213-0210
1110	Gloria Stevens	212-2211
1111	Victor Hugo	212-1661
✓1112	Paul Thomas	212-7898
1113	Neil Carter	214-5090
1114	Clara Barton	212-1231