A Guide to Patient Sampling

Survey results are only as good as the sample from which they came. Ideally, a practice could conduct a census where every patient would be asked to take a survey. In some rare instances, when a practice has a very small patient population, this may be possible. However, in most cases, there is seldom justification for population sampling given that they are expensive, time consuming, and will result in findings similar to those obtained from a smaller, valid sample.

Sampling has to be done correctly to ensure that it is valid. It is imperative that sampled patients be representative of the patient population. To ensure that your sample accurately represents your patient population, you must (a) clearly define the characteristics of your population, (b) choose the best method for selecting members of your sample, (c) determine the appropriate sample size, and (d) properly administer your survey questionnaire.

Defining the Characteristics of Your Population

The first thing you will do is identify the population being surveyed. For example, if you want to compare patients' perceptions across the physicians in the practice, you will need to define a sample large enough to allow analysis of responses for patients of each individual physician. Or, you may want to compare patients' perceptions based on how services are paid. Other characteristics could include age, gender, ethnicity, or the length of time someone has been a patient. If you decide to analyze based on certain characteristics, it is imperative that you include in the survey questions related to the appropriate demographic characteristic so that you can identify how certain groups of patients responded to a particular question.

Selecting Your Sample

The sample that is surveyed should be representative of the patient population you have identified. The best way to do this is using random sampling, which guarantees that any variations between the sample and the population are due to chance and not to selection bias. Random sampling can be done several ways: simple, stratified and systematic.

Non-stratified Simple Random Sample

A simple random sample is similar to pulling names from a hat; however, actually pulling names from a hat would not be the most practical approach. One popular way to select a simple random sample is to use an N^{th}-Name Select. To select an N^{th}-name sample, you divide the number of individuals in the population by the number of individuals you want to sample. For example, the practice has 5,000 patients and wants to distribute surveys to 500 patients. The N = 10 (5,000/500), which means you would select every 10th name on your patient list and administer the survey to these individuals.

Stratified Random Sample

If you decide there are subgroups for which you want to compare results, a simple random sample may not ensure that all subgroups are represented equally in the sample. For example, you want to draw a random sample of 400 patients to represent all of the patients that belong to three different insurance plans that reimburse for services at varying levels. You know that 45% use Plan A, 30% use Plan B and 25% use Plan C. If you picked the names of 400 patients out of a hat, there is a chance, for example, that more than 45% use

Plan A, or less than 25% use Plan C, etc. Using stratified random sampling, you separate patients into your subgroups, or strata, before randomly selecting the sample. You would then either select the same number from each strata (disproportionate stratified sampling) or different numbers based on the proportion of each group to the overall population (proportional stratified random sampling). Randomly selecting 40 patients from each plan would be disproportionate stratified sampling. Randomly selecting 54 patients from Plan A, 36 from Plan B and 30 from Plan C would be proportionate stratified sampling.

Non-stratified simple random samples and stratified random samples are good methods to use when you have a well-defined population and a complete list of the individuals in the population prior to selecting your sample. These methods would be used for quality assessment in a situation where data is collected a significant time after the patient encounters. However, many practices want to collect data on an ongoing basis without necessarily knowing or defining the patient population in advance, and want to administer the questionnaire during the time period immediately following or shortly after the point of care. In this scenario, systematic sampling is the best option.

Systematic Sampling

In systematic sampling, you initially select a patient from your identified patient population at random and distribute questionnaires to every nth patient in the identified patient population thereafter. The interval (n) is determined by dividing the number of patients seen during a similar survey period by the number of patients required for your sample. For example, your practice normally sees 1,000 patients during a one-month period of time and you would like to sample patients over a three-month period (3,000 patients). You have determined that you need to administer the survey to 500 patients to meet your sample size goal. In this case, n=6 (3,000 ÷ 500). You start by administering the survey to a patient upon his/her visit and then administer the survey to every sixth patient following until you have administered the survey to 500 patients.

Choosing the Right Sample Size

Your sample size is directly related to the accuracy of your survey results and how well they represent the perceptions of your population. For data to be useful, you will need to have a minimum sample size. Conversely, there is a point when increasing the size of your sample provides only negligible increases in the confidence of the data—in essence, a point of diminishing returns. For example, if you have a patient population of 3,000, collecting feedback from 50 of those patients would not provide you with reliable data that would be representative of all 3,000 patients. At the same time, collecting feedback from 350 randomly selected patients would provide you with reliable data that would not differ greatly from the feedback of 650 randomly selected patients or from all 3,000 patients. The trick is determining the minimum number of responses required to ensure your data is reliable and avoiding spending time and money collecting and analyzing more responses than you need.

Determining Your Level of Confidence

When calculating your sample size, you will first need to determine what will be your preferred confidence interval and confidence level. The confidence interval is a range of values of a sample statistic that is likely to contain a population parameter. In layman terms, it's the plus-or-minus figure usually reported in newspaper or television opinion poll results. For example, if you use a confidence interval of ±5% and 47% percent answer "yes" to a question, you can be "sure" that if you had asked the question of the entire relevant population, between 42% (47-5) and 52% (47+5) would have answered "yes." The confidence level represents how often the true percentage of the population who would pick an answer lies within the confidence interval. For example, the 95% confidence level means you can be 95% certain; the 99% confidence level means you can be 99% certain. Most market researchers use the 95% confidence level with a confidence interval of ±5%. If this were the case for this example, when you put the confidence level and the confidence interval together, you would say that you are 95% sure that between 42% and 52% of the entire population would have answered "yes."

Calculating Your Sample Size

The fastest and easiest way to calculate your sample size based on your desired confidence interval and confidence level is to use an electronic sample size calculator. There are many of these calculators available online free of charge. You would need only to type in the words "sample size calculator" into a search engine. For those individuals who cannot access an electronic sample size calculator, or who would like to manually calculate the size of their sample, there are many approaches and formulas. For convenience, a simple manual calculation based on proportions is provided.

$$n \geq \frac{(N)(p)(1-p)}{(N-1)(D) + (p)(1-p)}$$

$$D = \frac{(\text{Confidence Interval})^2}{Z^2}$$

n = Sample size
N = population size (the size of the population from which the sample will be selected)
P = Prior assumption, which is the proportion of respondents you would expect to answer a question a certain way. For example, if you have good reason to believe that 75% of respondents will likely respond "yes" to a question, your p would equal .75. In cases where you are not sure how respondents will respond, or you have multiple questions, you can take the most conservative approach and use .50, which translates into your respondents are just as likely to choice one answer over another.
Z = Area under normal curve corresponding to the desired confidence level. Following are the value of Z for three confidence levels: 1.645 (90%), 1.960 (95%), and 2.575 (99%).

Using the earlier example of a patient population of 3,000 and a desired 95% confidence level with a confidence interval of ±5%, the sample size would be calculated as follows.

$$D = \frac{(.05)^2}{(1.960)^2}$$

Or D = 0.00065

$$n \geq \frac{(3{,}000)(.5)(.5)}{(2{,}999)(0.00065) + (.5)(.5)}$$

Or n ≥ 341

Distributing the Questionnaire

One important factor is determining exactly how many surveys you will need to distribute in order to get the number of responses you will need for your survey sample. Different distribution methods can produce different response rates. For example, handing out surveys to patients in the office will often result in a response rate of 75% while mailing a survey to patients will most likely result in a 30-35% response rate. These are generalizations and can vary from practice to practice. Experience will be the teacher. If you have no prior experience surveying your patients and do not have a good idea of whether or not the response will be close to what is considered standard, be conservative the first time out of the gate and distribute the survey to a large enough group of patients so that you will attain your desired sample size even if the response rate is much lower than what was expected. For example, if you will be distributing a survey to patients in the office and you have no previous experience, calculate the number of surveys to be distributed based on a 65% response rate. If you need 350 individuals in your sample, distribute the survey to 540 patients.

Finally, when distributing the survey, particularly in situations where staff is handing out a survey or providing instructions for taking an online survey to every N^{th} patient, be careful to avoid any bias. Staff may be tempted to skip a patient who appeared to be particularly upset or has a reputation of being dissatisfied. Conversely, staff may be tempted to hand out a survey to a patient who appears particularly satisfied even if that patient is not designated as one of the patients to receive the survey according to your systematic sampling.

As you can see, there is a certain science to successful patient sampling. Follow these steps and you can feel confident that the data you collect will adequately represent the perceptions of your overall patient population.