We all know that size is important. Too small, and you may not be able to tell a difference; too big, and even minor changes may appear significant. While this provocative start may have gotten your attention, I must point out that I am, of course, talking about sample size in research.

When evaluating clinical research, we often think about the conclusions. As health care providers, we want to know whether or not the research is relevant to the patients that we treat every day. Caution is warranted when trying to make such decisions about relevance and applicability of research, because we must first think about the quality of the research and whether the results are valid. One aspect of evaluating quality is considering the sample size of a study.

Why is sample size important? Well, for one thing, it is important to determine if a study is adequately powered so that a difference, if a difference actually exists, can be detected between two groups, typically a test group and a control group. The power should be set at a minimum of 80%. This means that there is an 8 in 10 chance of detecting a difference between the groups if a difference actually exists. If our sample size is too small, then normal variation or randomness may impact the results, preventing us from seeing differences that may really exist.

Let’s consider an example. If we have a bag of M&Ms and randomly pick three pieces, by chance we could come up with three blue ones. But concluding from that sample of three that the bag has only blue M&Ms would be false, because we took too few pieces to make a meaningful assessment. In other words, the sample size was not large enough to be a representative sample of the entire bag of M&Ms, and our results could be easily skewed by chance.

If too small of a sample size is a problem, are there problems with a sample size that is too large? A larger sample size will provide results that are more strongly correlated with the general population and will also increase precision in the results. However, too large of a sample size can make very small differences that are clinically meaningless appear to be statistically significant.

Let’s go back to our M&M example. If we sampled 60,000 M&Ms that came from bags including all six colors—red, blue, green, brown, yellow, and orange—one would expect to find 10,000 pieces of each color if they were evenly distributed. Finding 9,950 red pieces, a difference of 0.5%, could lead to statistically significant results simply because so many M&Ms were sampled. But practically speaking, we would not perceive any meaningful difference in the number of red M&Ms. In other words, although the results might be statistically significant, they would not be practically (or clinically) relevant. This is more likely to happen in larger studies simply because of the larger sample size.

Clinical relevance, also called clinical significance, must not be overlooked even in studies that have an optimal sample size. Too often we jump to conclusions when seeing results that are statistically significant, typically having a P value < .05. A statistically significant result means that there is a difference...
between groups, but this has no indication on how big the difference is. If the difference between groups is very small, say a 0.5% difference in M&Ms or a 0.1-mm difference in probing depth, the result may be statistically significant, but the difference will be too small to be perceived by the patient or the clinician and therefore not clinically relevant. In an ideal world, we would like to see statistically significant results (ie, the treatment works) that are also clinically relevant (ie, the magnitude of the treatment effect is noticeable), but depending on sample size, these two calculations may not always correlate the way we would like them to.

Besides the scientific problems associated with having too small or too large of a sample size, there are also ethical and cost issues to address. A study with too small of a sample size to provide meaningful results might expose subjects to potential harm for no reason, and investigators would spend significant funds in conducting the study—so why even do the study in the first place? Having too large of a sample size may expose too many people to potential harm and will also certainly cost a significant amount of money to conduct. We are much better off from scientific, ethical, and economic standpoints when studies are conducted with the optimal number of subjects so that the results will be meaningful and useful.

Because sample size is so important and impacts the usability of the study results, ethics, and costs, it should always be calculated before a study is conducted. As stated above, the power should be set at a minimum of 80% for the calculation. If a sample size calculation is performed, this information is usually presented in the methods section, or possibly the results section, of a study.

So the bottom line is to be wary of studies that show no difference between groups when the sample size is small, as there may actually be differences that the study could not detect. Also be wary of studies that show very small differences between groups when the sample size is very large, as the differences may not be clinically relevant. Check the methods section to see if a sample size calculation was performed, and then check the results to see if the target sample size was met. Look for results that are both statistically significant and clinically relevant.

Julie Frantsve-Hawley, RDH, PhD
Editor-in-Chief

References