Statistical and Clinical Significance

In most quantitative research, various tests are used to answer the question of interest. When results are reported, the statistical test score is provided with the significance level. Most of the journal's readers probably understand that if a test is significant at whatever level was set, the results were positive in the predicted direction. In this column, I will discuss the meaning of the tests of significance as well as how statistical significance relates to clinical significance.

Statistical Significance

Statistical significance is the probability that an effect seen in a study is not likely to be due only to chance variation (Polit & Beck, 2010). The statistic is testing the null hypothesis, a statistical concept that no relationship exists between the two variables being studied. In reality, even unrelated variables will show some relationship or some difference in the case of experiments due to normal variation of data. So, the test is determining if the null hypothesis is probably true or false (Goodman, 2008). We assume if the null is determined to be false, then the research hypothesis concerning a relationship or difference is true. In other words, the test determines if the difference is one that might have occurred due to normal variation or if there is a likelihood of a relationship. Statistical significance is conveyed in probability terms (this is the source of the letter p). If the test is significant, there is a probability of a relationship. Significance level is used as a value judgment about the results of the test; however, it is not the same as clinical importance, which will be discussed later (Polit & Beck, 2010).

The test for significance is based on whatever statistical test was done in the study. For example, in both studies reported in this issue (Ambutas, Staffileno, & Fogg, 2014; Wrobleski, Joswiak, Dunn, Maxson, & Holland, 2014), researchers used t-tests and chi-squared tests to determine the differences between two groups and they reported the significance level for each of their results. With a significance level of $p<0.05$, we are assuming a 5% possibility the results were actually nonsignificant or a Type I error (Rempher & Miller, 2008). To ensure an even smaller chance of having a Type I error, the statistical significance test can be set at 0.01. The level of significance should be set prior to the analysis. Researchers should report the actual significance statistic, which is indicated by p (American Psychological Association, 2010).

Lynne M. Connelly, PhD, RN, is Associate Professor and Director of Nursing, Benedictine College, Atchison, KS. She is Research Editor for *MEDSURG Nursing.*

The choices of 0.05 or 0.01 for a significance level are based on convention and are arbitrary. There are many misconceptions about significance levels (Goodman, 2008). The significance of a particular test score should be considered only one piece of information about a study. In particular, significance levels do not indicate the magnitude of the relationship or how big the difference is between the control group and the experimental group (Conn, Chan, & Cooper, 2014). This leads to the key concept of clinical significance or clinical importance.

Clinical Significance

Readers should not confuse statistical significance with clinical significance. If there is a large enough sample, even a trivial relationship can be statistically significant. Even when there is a statistical significant relationship, however, there may be little clinical importance. In addition, the question arises of whether patient treatment should be changed based on the results (Man-Son-Hing et al., 2002). For example, as part of a study, patients took a test of their knowledge of diabetes, a class then was given, and they were tested after the class. If, on average, the patients answered one more question correctly on the test than a control group, is that really clinically significant? This is why confidence intervals are receiving more attention in medical and nursing research (Conn et al., 2014). Confidence intervals identify the range within which the real results would fail on 95% of the occasions, if the study was conducted many times. This adds to our understanding, but is still insufficient. Statistical tests alone cannot identify how important the results are to clinical practice.

Researchers and readers of research should consider what would make the results clinically meaningful. Clinical significance measures how large a difference in treatment must be in order for a change in practice to occur, or how large a relationship would be needed for significance. This requires judgment by clinicians. Wrobleski and colleagues (2014) addressed not only the significance level of the length-of-stay outcome, but also provided the average of the two groups (a difference of almost 3 days). A reduction in the length of stay of 3 days is likely to be a clinically important difference for patient care. In health care research, clinical significance can be determined by consensus of expert opinion or left to readers to determine for their own patient populations.

Clinical importance can be used with confidence intervals to determine an appropriate sample size. Ambutas and co-authors (2014) determined a sample size of 200 was necessary to allow detection of a 2% change in the rate of pressure ulcers between the two taping devices, as this would be

the lowest rate of improvement that would have an effect on patient satisfaction. This is an example of using clinical importance to make research decisions and to increase the power of the study to detect change. This is known as the *minimally detectable difference*, which helps quantify clinically important issues (Man-Son-Hing et al., 2002).

Another way to examine clinical significance is with the effect size. The magnitude or effect size of an intervention or relationship can be determined statistically. Small, medium, and large treatment effects can correlate with Cohen's *d* indices of 0.2, 0.5, or 0.8. These statistics can give readers an idea of how big the change or effect is when comparing one group to another. Other tests of treatment effects are relative risk ratios, odds ratios, number to treat, and area under the curve (McGough & Faraone, 2009).

Ultimately, clinical importance of a particular result must be determined by the clinician. Clinicians understand their patients, the clinical course of their conditions, and the practice setting. Readers should consider both statistical significance and clinical importance when examining the results of research studies. If more information is needed, a good statistical textbook or the references in this column may be helpful. MSN

REFERENCES

Ambutas, S., Staffileno, B.A., & Fogg, L. (2014). Reducing nasal pressure ulcers with an alternative taping device. *MEDSURG Nursing, 23*(2), 96-100.

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Conn, V.S., Chan, K.C., & Cooper, P.S. (2014). The problem with *p. Western Journal of Nursing, 36,* 291-293.

Goodman, S. (2008). A dirty dozen: Twelve p-value misconceptions. *Seminars in Hematology, 45*(3), 135-140

Man-Son-Hing, M., Laupacis, A., O'Rourke, K., Molnar, F.J., Mahon, J., Chan, K.B.Y., & Wells, G. (2002). Determination of the clinical importance of study results. *Journal of General Internal Medicine, 17*(6), 469-476.

McGough, J.J., & Faraone, S.V. (2009). Estimating the size of treatment effects. *Psychiatry, 6*(10), 21-29.

Polit, D.F., & Beck, C.T. (2010). *Essentials of nursing research: Appraising the evidence for nursing practice*. Philadelphia, PA: Wolters Kluwer/Lippincott Williams & Wilkins.

Rempher, K.J., & Miller, S. (2008). Making sense of statistical power. *American Nurse Today, 3*(6), 36-38.

Wrobleski, D.M.S., Joswiak, M.E., Dunn, D.F., Maxon, P.M., & Holland, D.E. (2014). Discharge planning rounds to the bedside: A patient- and family-centered approach. *MEDSURG Nursing, 23*(2), 111-116.