A practitioner's guide to developing critical appraisal skills

What is the difference between clinical and statistical significance?

Romina Brignardello-Petersen, DDS, MSc; Alonso Carrasco-Labra, DDS, MSc; Prakeshkumar Shah, MSc, MBBS, MD, DCH, MRCP, FRCP(C); Amir Azarpazhooh, DDS, MSc, PhD

Investigators in a study published in 2010 compared the efficacy of nimesulide with that of meloxicam (two nonsteroidal anti-inflammatory drugs) in the control of postoperative pain, swelling and trismus after extraction of impacted mandibular third molars.[1] Among their conclusions, the authors stated that "[nimesulide] was more effective than [meloxicam] in the control of swelling and trismus following the removal of impacted lower third molars."[1] This conclusion was supported by the results observed in their randomized clinical trial. The authors reported that after the third molar surgical extraction, patients experienced a reduction in mouth opening, but that this reduction was significantly larger at 72 hours after surgery when patients had received meloxicam than when patients had received nimesulide. The

ABSTRACT

JADA CE CONTINUING EDUCATION ARTICLE 1

Background. It is common to find published studies in which the authors claim to have found significant results. However, many times these results are only statistically significant with no meaningful impact in clinical settings.

Methods. The authors aim to clarify and differentiate the concepts of statistical and clinical significance, as well as to provide guidance on how to interpret research results to determine whether an observed difference is meaningful.

Results. Study results are considered to be statistically significant if statistical tests that examine the null hypothesis of no difference yield *P* values that are smaller than the significance level prespecified by the authors. In this way, researchers can use hypothesis testing to assess the possibility that observed results could have arisen by chance. However, hypothesis testing cannot establish the clinical implications of these results. Rather, clinical significance can be established once the magnitude of results is larger than the minimal clinically important difference. Clinical significance then would encompass not only statistical significance, but also the importance of the outcomes to patients, clinicians and policymakers.

Conclusion. The values for statistical significance alone cannot convey the complete picture of the effectiveness of an intervention or of a difference between two groups. Both clinical and statistical significance are important measures for interpretation of clinical research results and should complement each other.

Practical Implications. Any benefit in terms of improved health outcomes must be both clinically and statistically significant. If there is no benefit at the threshold of both clinical and statistical improvement, then the intervention should not be used for that purpose.

Key Words. Statistics; epidemiology; decision making; statistical significance; clinical significance.

JADA 2013;144(7):780-786.

Dr. Brignardello-Petersen is a lecturer, Evidence-Based Dentistry Unit, Faculty of Dentistry, University of Chile, Santiago.
Dr. Carrasco-Labra is an instructor, Evidence-Based Dentistry Unit, Faculty of Dentistry, University of Chile, Santiago.
Dr. Shah is an associate professor, Clinical Epidemiology Program, Institute of Health Policy, Management and Evaluation, University of Toronto; an associate professor, Department of Paediatrics, Faculty of Medicine, University of Toronto; and a staff neonatologist, Mount Sinai Hospital, Institute of Health Policy, Toronto.
Dr. Azarpazhooh is an assistant professor, Disciplines of Endodontics and Dental Public Health, Faculty of Dentistry, University of Toronto; an assistant professor, Clinical Epidemiology Program, Institute of Health Policy, Management and Evaluation, Faculty of Medicine, University of Toronto; and a collaborator, Toronto Health Economics and Technology Assessment Collaborative, University of Toronto. Address reprint requests to Dr. Azarpazhooh, Dental Research Institute, Faculty of Dentistry, University of Toronto, Room 515-C, 124 Edward St., Toronto, Ontario, M5G 1G6, Canada, e-mail amir.azarpazhooh@dentistry.utoronto.ca.

authors reported a *P* value of .03 for the difference in the mean reduced mouth opening of 1.39 centimeters in the nimesulide group versus 1.7 cm in the meloxicam group. This difference of 3.1 millimeters was the basis for the authors' claim of the superiority of nimesulide. However, from a clinical perspective, this difference does not seem large. How can we know if these numbers show that the reduction in mouth opening is significantly larger when patients received meloxicam therapy, as the authors report? What do the authors mean when they use the expression "significantly larger"? Is a *P* value < .05 sufficient to claim that there is a significant difference?

In this article, we aim to clarify and differentiate the concepts of statistical significance and clinical significance, as well as to provide guidance on how to interpret research results to determine whether an observed difference is clinically meaningful.

STATISTICAL SIGNIFICANCE

It is not feasible to conduct a study in which investigators study all potential patients. Thus, researchers have to base their conclusions on a sample of people and then determine the probability or likelihood that a conclusion made on the basis of an analysis of data from this sample will hold true when applied to the population as a whole.[2]

Researchers have used statistical significance for many years as a means to assess the effects of interventions in clinical research and to show that observed differences likely are not due to chance.[3] Usually, the claim of statistical significance depends on obtaining a specific *P* value after conducting a statistical significance test, as in the earlier example.

A *P* value is the probability of obtaining a mean difference that is at least as far from a specified value (null value) as the mean observed in the study, given that this specified value is the true value.[4] In the example above, if we assume that the true difference in mouth-opening reduction between nimesulide and meloxicam is 0 mm, what the authors found was a 3 percent probability of observing the 3.1-mm difference (or larger) that they detected. Because the probability of that happening is so small, it is unlikely that the differences they observed were due to chance; thus, they could claim that there are real and statistically significant differences between the two treatments.

As stated earlier, the *P* value is obtained when conducting statistical hypothesis testing. To perform this test, we start by assuming that the result of interest (the mean or proportion of the outcome of the study) is equal to some specific value. This claim is called the null hypothesis. In the example, the null hypothesis was that there is no difference in mouth-opening reduction between the two drug groups. The investigators then construct an alternative hypothesis such that it contradicts the null hypothesis. In this case, the alternative hypothesis was that differences existed between the drugs with regard to mouth-opening reduction.[5] The next step is to compare the data obtained in the study with the value specified in the null hypothesis—using the probability theory—to attain a *P* value. The *P* value is related to how much the data contradict the null hypothesis. If a large *P* value is obtained, the data are consistent with the null hypothesis. Conversely, if a small *P* value is obtained, the data contradict the null hypothesis, and the results are unlikely to have occurred if the null hypothesis actually were true. However, the investigators must decide whether the *P* value is sufficiently small to reject the null hypothesis. Although it is arbitrary, a *P* value of .05 has been the conventionally accepted value for level of significance.[6]

Type I error. The level of significance reflects the probability of committing a type I error—that is, rejecting the null hypothesis when it actually is true.[7] In other words, it is the probability of falsely claiming that there is a difference in mouth-opening reduction when there is not. According to the earlier description, the *P* value is not the probability that the null hypothesis is true. This is a common misconception. A large *P* value does not mean that the null hypothesis is true; at best, it implies that the study results are inconclusive. Likewise, a small *P* value does not mean that the alternative hypothesis is true; at best, it implies that the data are incompatible with the null hypothesis' being true.[5]

Type II error. On the other hand, a probability exists of not rejecting the null hypothesis when it is false, which is known as a type II error. A type II error occurs when researchers fail to observe a difference between interventions even though a true difference does exist.[8] For example, imagine a study in which the researcher wants to determine whether the incidence of cleft lip and palate is larger in one of two towns. Let us assume that a difference between the towns truly exists, and that the true incidence in town A is five in 1,000 newborns, whereas in town B, it is one in 1,000 newborns. If the

ABBREVIATION KEY. MCID: Minimal clinically important difference.