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

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.

TABLE 1

Relationship between sample size and *P* values (assuming constant means and SDs*).†

HYPOTHETICAL SCENARIO	MEAN (SD) REDUCTION IN MOUTH OPENING AT 72 HOURS, IN CENTIMETERS		SAMPLE SIZE	*P* VALUE‡	STATISTICAL SIGNIFICANCE
	Meloxicam	Nimesulide			
1	1.70 (0.6)	1.39 (0.75)	20	.321	No
2	1.70 (0.6)	1.39 (0.75)	40	.157	No
3	1.70 (0.6)	1.39 (0.75)	60	.082	No
4	1.70 (0.6)	1.39 (0.75)	80	.045	Yes
5	1.70 (0.6)	1.39 (0.75)	100	.025	Yes
6	1.70 (0.6)	1.39 (0.75)	120	.014	Yes

* SDs: Standard deviations.
† This table illustrates the influence that sample size has on *P* values. When the mean of the outcome and its SD are constant, an increase in sample size leads to smaller *P* values.
‡ Two-sided unpaired *t* test; significance level ≤ .05. *P* values were calculated by using statistical software (The R Project for Statistical Computing, The R Foundation for Statistical Computing, Vienna).

researcher observes only 50 newborns in each town, it is likely that he or she will not find an infant with cleft lip and palate in either town. At the end of the study, the data will suggest that the incidence of this malformation is 0 for both towns. Therefore, the researcher will claim falsely that no difference exists between the two towns with regard to the incidence of cleft lip and palate, because he or she failed to find any infant with the malformation. This is an issue of "power" of the study.

Study power. The power of a study is its capacity to detect differences that truly exist, and it is defined as the probability of rejecting the null hypothesis when it is false.[5] Power is the opposite of a type II error; higher power implies a smaller probability of committing a type II error, and vice versa. Thus, our hypothetical study of the incidence of cleft lip and palate in two towns was underpowered. Also, as this example illustrates, the power of a study depends, in part, on the sample size (the number of newborns observed) and the effect size (the difference in the incidence of the malformation between the groups).

ISSUES PERTAINING TO STATISTICAL SIGNIFICANCE

Understanding statistical significance requires thinking in terms of probability. At the completion of a statistical significance test, there are two possible outcomes: reject the null hypothesis or fail to reject the null hypothesis. This qualitative result often is used as a substitute for quantitative scientific evidence.[7] As illustrated in Tables 1 and 2, one issue of concern is that the results of statistical testing are influenced highly by the sample size and variability within the sample. Table 1 shows that increasing the sample size, while keeping everything else constant, results in a smaller *P* value's having been obtained in hypothesis testing. This leads to statistically significant results when the sample size is larger and to nonstatistically significant results when the sample size is smaller. Table 2 shows that, while keeping everything else constant, a smaller variation in the response to an intervention among participants in one group results in a smaller *P* value's having been obtained in statistical testing. Consequently, statistically significant results are obtained when the variability is smaller, and nonstatistically significant results are obtained when the variability is larger.

Therefore, studies in which the sample size is large, in which there is little variability within the sample or both are more likely to lead to statistically significant results compared with identical studies in which the sample sizes are smaller and the variability is greater. This is true even when the effect size (the difference between the groups) is the same, as shown in Tables 1 and 2.

In the study in which De Menezes and Cury[1] compared the efficacy of nimesulide versus that of meloxicam for the control of postoperative pain, swelling and trismus after extraction of impacted mandibular third molars,[1] let us imagine that the authors had measured the outcome pain as dichotomous (that is, the presence or absence of pain). In addition, let us suppose that after completing the trial, these authors observed that 85 percent of participants who received nimesulide therapy reported experiencing no postoperative pain, whereas 90 percent of participants allocated to the meloxicam group reported experiencing no postoperative pain. If we compare these two proportions in a trial in which researchers enrolled a total of 80 pa-

tients, the *P* value for the hypothesis testing would be .25. On the other hand, had researchers in the trial enrolled 800 participants, the *P* value for the same comparison would be .016. Although the difference in the proportion of patients with no postoperative pain in both trials was 5 percent, the conclusion drawn in the trial with the smaller sample size is that there was no statistically significant difference between the two drugs, whereas in the trial with the larger sample size, the authors could have made the opposite claim.

TABLE 2

Relationship between variability estimates and *P* values (assuming constant means and sample size).*

HYPOTHETICAL SCENARIO	MEAN (SD†) REDUCTION IN MOUTH OPENING AT 72 HOURS, IN CENTIMETERS		SAMPLE SIZE	*P* VALUE‡	STATISTICAL SIGNIFICANCE
	Meloxicam	Nimesulide			
1	1.70 (0.40)	1.39 (0.55)	60	.015	Yes
2	1.70 (0.45)	1.39 (0.60)	60	.027	Yes
3	1.70 (0.50)	1.39 (0.65)	60	.043	Yes
4	1.70 (0.55)	1.39 (0.70)	60	.061	No
5	1.70 (0.60)	1.39 (0.75)	60	.082	No
6	1.70 (0.65)	1.39 (0.80)	60	.105	No

* This table illustrates the influence that sample variability has on *P* values. When the mean outcome value and sample size are constant, an increase in sample variability leads to higher *P* values.
† SD: Standard deviation.
‡ Two-sided unpaired *t* test; significance level ≤ .05. *P* values were calculated by using statistical software (The R Project for Statistical Computing, The R Foundation for Statistical Computing, Vienna).

In light of the above, the main problem is that hypothesis testing can omit any differences that were observed in the study. Statistical testing indicates only the probability of the observed differences'—without regard to the size of the differences—occurring by chance.[2] By using the hypothesis testing approach and claiming that there are differences in effects of interventions on the basis of a *P* value alone, we lose valuable information regarding the size of the effect.[9] This is illustrated in Table 3, which shows different effect sizes leading to the same conclusion: the results are or are not statistically significant.

Researchers can reach the same *P* value in many ways by combining different treatment effects, within-group variability and sample sizes. Thus, the results of a statistical test cannot indicate whether a treatment effect is important enough to be useful for patients.[10] Moreover, some readers may erroneously interpret small *P* values as large effects,[11] when, in fact, the *P* values actually represent only the probability of having observed what was observed in the study and indicate nothing about the effect size.

CLINICAL SIGNIFICANCE

In 1984, Jacobson and colleagues[12] proposed the term "clinical significance" as a means of evaluating the practical value of a treatment. Although the literature contains many definitions of and discussions about this term,[9,13-18] most authors agree that a clinically significant result must fulfill the following criteria:

- A change in an outcome or a difference in outcome between groups occurs that is of interest to someone; patients, physicians or other parties interested in patient care conclude that the effect of one treatment compared with another makes a difference.
- The change or difference between groups must occur in an important outcome. It can be any outcome that may alter a clinician's decisions regarding treatment of a patient, such as a reduction in symptoms, improvement in quality of life, treatment effect duration, adverse effects, cost effectiveness or implementation.
- The change or difference must be statistically significant. The difference must be greater than what may be explained by a chance occurrence.[9,13-18]

Minimal clinically important difference. Jaeschke and colleagues[19] introduced the concept of "minimal clinically important difference" (MCID) to determine whether a difference between treatments is of interest. They defined the MCID as "the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive costs, a change in the patient's management." Even though this definition reflects the patient's perspective, it can be applied readily to any party involved in the health care chain (such as clinicians, family members, policymakers, hospital administrators).

Although each member of the health care

TABLE 3

Relationship between effect size increase and statistical significance (assuming constant mean in meloxicam group and constant standard deviations in both groups).*

HYPOTHETICAL SCENARIO	MEAN (SD†) REDUCTION IN MOUTH OPENING AT 72 HOURS, IN CENTIMETERS		DIFFERENCE IN MEAN REDUCTION IN MOUTH OPENING, IN CENTIMETERS	SAMPLE SIZE	STATISTICAL SIGNIFICANCE‡
	Meloxicam	Nimesulide			
1	1.70 (0.6)	1.39 (0.75)	0.31	40	No
2	1.70 (0.6)	1.34 (0.75)	0.36	40	No
3	1.70 (0.6)	1.29 (0.75)	0.41	40	No
4	1.70 (0.6)	1.24 (0.75)	0.46	40	Yes
5	1.70 (0.6)	1.19 (0.75)	0.51	40	Yes
6	1.70 (0.6)	1.14 (0.75)	0.56	40	Yes

* This table illustrates that when statistical hypothesis testing is used, the conclusion drawn (that is, whether differences are or are not statistically significant) does not reflect how much larger the effect of one intervention is compared with another. Even though the mean difference in outcomes across the scenarios is increasing, the conclusion derived from statistical hypothesis testing is exactly the same.
† SD: Standard deviation.
‡ Two-sided unpaired *t* test; significance level ≤ .05. *P* values were calculated by using statistical software (The R Project for Statistical Computing, The R Foundation for Statistical Computing, Vienna).

team has his or her opinion as to what constitutes an important difference, the patient's perspective should take priority.[10] Because this is a relevant topic for interpreting the results of clinical studies, investigators have made many efforts to provide guidance and to determine the MCID for outcomes in many clinical areas, including dentistry.[20-25]

Perceptions regarding which outcomes are important likely will vary within and across cultures, as well as on the basis of the perspective being considered (such as the patient's, the clinician's).[26-28] An important outcome from the patient's perspective has been defined as one in which the patient, if he or she knew the outcome would be the only thing that would change with a treatment, would consider receiving this treatment even if it is associated with adverse effects, inconvenience or cost.[29] For example, most patients with oral cancer consider mortality to be an important outcome, and many are willing to receive chemotherapy despite its adverse effects to reduce the chance of dying. However, other perspectives often are important as well. Thus, the relative importance given to outcomes should reflect the perspectives of those affected.[30]

Although little evidence exists to support any judgment about the outcomes that are important in dentistry, some guidance is found in other medical fields. For example, when ranking outcomes for assessing the effectiveness of a drug for treating renal failure, most patients would agree that mortality is the most critical outcome, whereas an adverse effect such as flatulence is of lower importance.[30] In addition, with regard to surrogate outcomes, researchers generally agree that because the evidence is associated with a surrogate measure, it is weaker and, thus, of lower importance.[30,31]

Surrogate outcomes. According to Temple,[32] a surrogate endpoint in the context of a clinical trial

> is a laboratory measurement or a physical sign used as a substitute for a clinically meaningful endpoint that measures directly how a patient feels, functions or survives. Changes induced by a therapy on a surrogate endpoint are expected to reflect changes in a clinically meaningful endpoint.

In other words, a surrogate outcome is an outcome that is measured in place of a biologically definitive or clinically most meaningful outcome.[33] An example of a potentially misleading surrogate outcome in dentistry is a significant reduction in salivary mutans streptococci as a substitute for the occurrence of caries lesions,[34-36] or measuring pocket depth as a surrogate outcome for tooth loss or for a patient's quality of life.[34] Other examples of potentially misleading surrogate outcomes are biomarkers that have not been validated, such as the use of components of gingival crevicular fluid[37] or venous blood,[38] which are ubiquitous in clinical trials in periodontics.[39] Such surrogate outcomes and biomarkers are trustworthy when they meet two criteria: there is a strong, independent and consistent association between the surrogate measure and the clinically meaningful outcome it is trying to measure; and the evidence shows improvement in both the clinically meaningful outcome and the surrogate outcome. Because this does not happen frequently, surrogate outcomes are not considered important to patients.[40]

INTERPRETING RESEARCH FINDINGS

To determine whether the results of a study are clinically significant, we need to consider the three criteria described earlier. In addition, we must take into consideration the fact that the results obtained in a study reflect the specific sample used in the study. Consequently, even if the study were replicated, the subsequent *P* value would be different from the first.

Confidence intervals. The use of confidence intervals (CIs) helps researchers and clinicians interpret both statistical and clinical significance and is a way to estimate what the results would be in the population.[41] A CI can be defined as a range of values calculated from the data within which the investigator believes a true parameter value will lie with some specific probability.[42] In other words, a CI is a plausible range of values within which the true value actually will lie given the data observed in a study.[43] For example, consider a hypothetical study in which investigators aim to estimate the percentage of successful short versus conventional dental implants. The study results show a 10 percent difference in favor of the conventional implants. The 95 percent CI is 2 to 18 percent, which means that we can be 95 percent confident that the true (but unknown) difference between the two types of implants is between 2 and 18 percent. A study's CI either includes the true population difference or it does not. Thus, CIs provide information about the magnitude of an effect and the uncertainty surrounding it, which may help us to evaluate the clinical significance of the effect.[41,44]

CIs and MCID. Using the CI and the concept of MCID, we find that a difference observed in a study will be clinically significant when its 95 percent CI is higher than the MCID. If the 95 percent CI contains the MCID, then no conclusions regarding the clinical significance of an effect can be made.[44] However, readers must be careful when using CIs to assess clinical significance, because the width of a CI is associated with the sample size. Let us suppose that the MCID for the example above is 4 percent. Because the CI indicates that the true difference between the implant types likely will be between 2 and 18 percent, we cannot claim that the study results are clinically significant. However, if the sample size were tripled, the 95 percent CI of the difference would be 5 to 15 percent. This range would indicate that it is not plausible for the difference between the two implant types to be as small as 4 percent, making it a clinically significant difference.

DISCUSSION

We have argued that 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 treatment should not be used for that purpose.

In their study of the effects of nimesulide versus meloxicam on trismus after tooth extraction, De Menezes and Cury[1] reported that a 3.1-mm difference between the groups at 72 hours was statistically significantly large. As discussed earlier, the statistical testing process they used to compare the two interventions correctly supports this claim. However, their study lacked a deeper examination of the issue of significance.

To reiterate, the three main characteristics of a clinically significant effect are that the change (that is, the difference between the results of the experimental group and those of the control group) is of interest to someone; the change occurred in an important outcome; and the change reached statistical significance. Limitations in mouth opening can have many negative consequences, such as problems in nutrition and speech. For healthy patients undergoing third molar extraction, it probably is an important complication to consider. Thus, the study finding by De Menezes and Cury[1] seems to fulfill two of the three criteria to be considered clinically important. However, to claim that a difference of 3.1 mm in trismus is of interest to clinicians or patients—which would meet the first criterion and make this result clinically significant—is questionable. Would a patient be willing to overcome adverse effects or bear higher costs to experience 3.1 mm less trismus? Whether the difference between two groups with regard to an observed outcome is of interest is a matter of debate in many cases. This reminds us of the major role that judgment gained through clinical experience plays in interpreting and applying research results.

CONCLUSION

It is important to realize that values for statistical significance alone cannot convey the complete picture of effectiveness of an intervention or a difference between two groups. Both clinical and statistical significance are important for interpretation of clinical research results and should complement each other. ■

Disclosure. None of the authors reported any disclosures.

1. De Menezes SA, Cury PR. Efficacy of nimesulide versus meloxicam in the control of pain, swelling and trismus following extraction of impacted lower third molar (published online ahead of print April 21, 2010). Int J Oral Maxillofac Surg 2010;39(6):580-584.

doi:10.1016/j.ijom.2010.03.012.
2. Norman G, Streiner D. Elements of statistical inference. In: Norman G, Streiner D, eds. Biostatistics: The Bare Essentials. 3rd ed. Hamilton, Ontario, Canada: B.C. Decker; 2008:46-62.
3. Bhardwaj SS, Camacho F, Derrow A, Fleischer AB Jr, Feldman SR. Statistical significance and clinical relevance: the importance of power in clinical trials in dermatology. Arch Dermatol 2004;140(12):1520-1523.
4. Greenland S, Rothman K. Fundamentals of epidemiologic data analysis. In: Rothman K, Greenland S, Lash T, eds. Modern Epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008:213-237.
5. Pagano M, Gauvreau K. Hypothesis testing. In: Pagano M, ed. Principles of Biostatistics. 2nd ed. Pacific Grove, Calif.: Duxbury; 2000:232-258.
6. Gauvreau K, Pagano M. Why 5%? Nutrition 1994;10(1):93-94.
7. Hayat MJ. Understanding statistical significance. Nurs Res 2010;59(3):219-223.
8. Koretz RL. Is statistical significance always significant? Nutr Clin Pract 2005;20(3):303-307.
9. LeFort SM. The statistical versus clinical significance debate. Image J Nurs Sch 1993;25(1):57-62.
10. Sackett DL. The tactics of performing therapeutic trials. In: Haynes RB, ed. Clinical Epidemiology: How to Do Clinical Practice Research. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2006:107.
11. Cleophas TJ. Clinical trials: renewed attention to the interpretation of the *P* values—review. Am J Ther 2004;11(4):317-322.
12. Jacobson NS, Follette WC, Revenstorf D. Psychotherapy outcome research: methods for reporting variability and evaluating clinical significance. Behav Therapy 1984;15(4):336-352.
13. Hujoel PP, Armitage GC, Garcia RI. A perspective on clinical significance. J Periodontol 2000;71(9):1515-1518.
14. Killoy WJ. The clinical significance of local chemotherapies. J Clin Periodontol 2002;29(suppl 2):22-29.
15. Kingman A. Statistical vs clinical significance in product testing: can they be designed to satisfy equivalence? J Public Health Dent 1992;52(6):353-360.
16. Lindgren BR, Wielinski CL, Finkelstein SM, Warwick WJ. Contrasting clinical and statistical significance within the research setting. Pediatr Pulmonol 1993;16(6):336-340.
17. Greenstein G. Clinical versus statistical significance as they relate to the efficacy of periodontal therapy. JADA 2003;134(5):583-591.
18. Hollon S, Flick S. On the meaning of clinical significance. Behav Assess 1988;10:10.
19. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. Control Clin Trials 1989;10(4):407-415.
20. Greenstein G, Lamster I. Efficacy of periodontal therapy: statistical versus clinical significance. J Periodontol 2000;71(4):657-662.
21. Allen PF, O'Sullivan M, Locker D. Determining the minimally important difference for the Oral Health Impact Profile-20 (published online ahead of print March 16, 2009). Eur J Oral Sci 2009;117(2):129-134. doi:10.1111/j.1600.0722.2009.00610.x.
22. Ingram M, Choi YH, Chiu CY, et al. Use of the minimal clinically important difference (MCID) for evaluating treatment outcomes with TMJMD patients: a preliminary study (published online ahead of print Jan. 10, 2012). J Appl Biobehav Res 2011;16(3-4):148-166. doi:10.1111/j.1751-9861.2011.00068.x.
23. Thomson WM. Measuring change in dry-mouth symptoms over time using the Xerostomia Inventory. Gerodontology 2007;24(1):30-35.
24. Tsakos G, Allen PF, Steele JG, Locker D. Interpreting oral health-related quality of life data (published online ahead of print Nov. 10, 2011). Community Dent Oral Epidemiol 2012;40(3):193-200. doi:10.1111/j.1600-0528.2011.00651.x.
25. Tsakos G, Bernabe E, D'Aiuto F, et al. Assessing the minimally important difference in the oral impact on daily performances index in patients treated for periodontitis (published online ahead of print June 1, 2010). J Clin Periodontol 2010;37(10):903-909. doi:10.1111/j.1600-051X.2010.01583.x.
26. Blackhall LJ, Murphy ST, Frank G, Michel V, Azen S. Ethnicity and attitudes toward patient autonomy. JAMA 1995;274(10):820-825.
27. Klessig J. The effect of values and culture on life-support decisions. West J Med 1992;157(3):316-322.
28. Ruhnke GW, Wilson SR, Akamatsu T, et al. Ethical decision making and patient autonomy: a comparison of physicians and patients in Japan and the United States. Chest 2000;118(4):1172-1182.
29. Akl EA, Briel M, You JJ, et al. LOST to follow-up Information in Trials (LOST-IT): a protocol on the potential impact. Trials 2009;10:40.
30. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 2—framing the question and deciding on important outcomes (published online ahead of print Dec. 30, 2010). J Clin Epidemiol 2011;64(4):395-400. doi:10.1016/j.clinepi.2010.09.012.
31. Schunemann HJ, Oxman AD, Fretheim A. Improving the use of research evidence in guideline development: 6—determining which outcomes are important. Health Res Policy Syst 2006;4:18.
32. Temple RJ. A Regulatory Authority's Opinion about Surrogate Endpoints. New York City: John Wiley & Sons; 1995.
33. Piantadosi S. Objectives and outcomes. In: Clinical Trials: A Methodologic Perspective. 2nd ed. Hoboken, N.J.: Wiley-Interscience; 2005:187-210.
34. Hujoel PP. Endpoints in periodontal trials: the need for an evidence-based research approach. Periodontol 2000 2004;36:196-204.
35. Forgie AH, Paterson M, Pine CM, Pitts NB, Nugent ZJ. A randomised controlled trial of the caries-preventive efficacy of a chlorhexidine-containing varnish in high-caries-risk adolescents. Caries Res 2000;34(5):432-439.
36. Sandham HJ, Brown J, Chan KH, Phillips HI, Burgess RC, Stokl AJ. Clinical trial in adults of an antimicrobial varnish for reducing mutans streptococci. J Dent Res 1991;70(11):1401-1408.
37. Oringer RJ, Al-Shammari KF, Aldredge WA, et al. Effect of locally delivered minocycline microspheres on markers of bone resorption. J Periodontol 2002;73(8):835-842.
38. Vidal F, Figueredo CM, Cordovil I, Fischer RG. Periodontal therapy reduces plasma levels of interleukin-6, C-reactive protein, and fibrinogen in patients with severe periodontitis and refractory arterial hypertension. J Periodontol 2009;80(5):786-791.
39. Pihlstrom BL, Barnett ML. Design, operation, and interpretation of clinical trials (published online ahead of print June 25, 2010). J Dent Res 2010;89(8):759-772. doi:10.1177/0022034510374737.
40. Bucher HC, Kunz R, Cook D, Holbrook A, Guyatt GH. Surrogate outcomes. In: Guyatt GH, Rennie D, Meade MO, Cook D, eds. Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice. 2nd ed. New York City: McGraw Hill; 2008:317-340.
41. Fethney J. Statistical and clinical significance, and how to use confidence intervals to help interpret both (published online ahead of print March 29, 2010). Aust Crit Care 2010;23(2):93-97. doi:10.1016/j.aucc.2010.03.001.
42. Piantadosi S. Notation and terminology. In: Piantadosi S, ed. Clinical Trials: A Methodologic Perspective. 2nd ed. Hoboken, N.J.: Wiley-Interscience; 2005:569-585.
43. Guyatt GH, Walter S, Cook D, Wyer P, Jaeschke R. Confidence intervals. In: Guyatt GH, Rennie D, Meade MO, Cook D, eds. Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice. 2nd ed. New York City: McGraw Hill; 2008:99-107.
44. Braitman LE. Confidence intervals assess both clinical significance and statistical significance. Ann Intern Med 1991;114(6):515-517.