A practical approach to evidence-based dentistry: X

How to avoid being misled by clinical studies’ results in dentistry

Alonso Carrasco-Labra, DDS, MSc, PhD(c); Romina Brignardello-Petersen, DDS, MSc; Amir Azarpazhooh, DDS, MSc, PhD, FRCDC(c); Michael Glick, DMD; Gordon H. Guyatt, MD, MSc

TENTH IN A SERIES

In previous articles in this series, we presented the process and main principles of evidence-based dentistry (EBD),1 how to search for evidence,2 and how to use articles about therapy,3 harm,4 diagnosis,5 systematic reviews,6 clinical practice guidelines,7 qualitative studies,8 and economic evaluations.9 In this final article of the EBD series, we offer clinicians guidance on how to avoid being misled by biased interpretations of study results.

Academic competition and conflict of interest have fueled misleading presentations of research findings published in peer-reviewed journals. Irrespective of whether a researcher works in academia or in the pharmaceutical industry, there is always a personal interest and a rising pressure to succeed and to provide novel and exciting findings; this pressure often results in interpretations of findings that are far more enthusiastic than the data warrant.10

In the area of psychopharmacology, for example, the investigators of 90% to 98% of industry-funded primary studies comparing 2 drugs reported results that favored the drug produced by their company, particularly when the active comparator drug was a rival product.11 This situation is not exclusive to primary studies. The investigators of industry-sponsored systematic reviews are less transparent regarding their methods, are less rigorous in their risk of bias assessment, and provide more favorable conclusions toward the study sponsor’s drug than are the investigators of reviews that have not been funded by the investigators’ industry.12 When companies employ ghostwriters to produce manuscripts under the names of credible and often well-known researchers, the reported results are likely to be overly favorable.13

ABSTRACT

Background and Overview. Clinicians using evidence to inform decisions on a daily basis have access to a number of tools to help them judge the importance of discriminating studies conducted using suboptimal methods from more rigorous ones. Many checklists have been developed to facilitate and guide clinicians to identify and critically appraise clinical studies. However, only limited guidance is available addressing how clinicians can identify misleading claims from those that can be supported reliably by study results.

Practical Implications. In this final article of a series of 10, the authors provide key concepts that clinicians can use to help them avoid using biased inferences or statements that are “too good to be true.”

Key Words. Results interpretation; misleading presentation of results; evidence-based dentistry.

JADA 2015:146(12):919-924

http://dx.doi.org/10.1016/j.adaj.2015.08.008

The involvement of members of a specific industry is not necessary for overenthusiastic interpretations of results. Academic investigators also are subject to the global industry of producing research evidence. The reward system in science involves receiving grants and having research results published, and scientists may believe that overplaying the significance of their work is a requirement for success.14

Although guidance and tools for clinicians to recognize study results that have a high risk of bias are widely available,15,16 researchers have made limited efforts to facilitate the identification of distorted interpretations and misleading presentations of the results of clinical studies. We present the following examples not to criticize investigators, but to illustrate the need to increase awareness among clinicians and encourage them to avoid putting excessive trust in investigators’ interpretations of their findings.

GUIDANCE ON HOW TO AVOID BEING MISLED BY THE RESULTS OF CLINICAL STUDIES

We present 7 criteria that dental professionals can follow to avoid being mislead by the results of clinical
2. Read synoptic abstracts published in secondary publications (preappraised resources) for evidence-based dentistry. Busy clinicians interested in using evidence to inform their clinical practice may not have time to skip the discussion sections of articles and instead critically appraise the evidence, and thus make sense of the results, by themselves. Secondary journals and sources, such as Evidence-Based Dentistry, Journal of Evidence-Based Dental Practice, and the American Dental Association’s Evidence Database, publish synoptic summaries in an abstract format that are accompanied by a brief summary of the original article and a critical appraisal conducted by a team of clinicians and methodologists. These abstracts, developed by independent third parties who have no conflicts of interest, reduce the distortion that the authors of a primary or secondary study may have introduced in the original article. Another objective of this type of synopsis is to educate clinicians about the methodological aspects of different study designs, thereby increasing clinicians’ critical appraisal skills.

BOX 1

Example: Is an herbal mouthrinse effective to reduce the dental biofilm and the incidence of caries?

The investigators of a crossover randomized controlled trial conducted in 12 healthy participants compared 2 mouthrinses commonly available in the market: a special experimental formula with extract and essential oil of Baccharis dracunculifolia and a control mouthrinse based on a basic formulation that did not contain an active component. The investigators followed the participants for 1 week and measured the mean values of biofilm. In the results section of the article, the authors did not provide any numerical data and only referred to the fact that differences between the groups were not statistically significant. The investigators’ failure to show differences between the treatments resulted in no evidence provided to support the intervention. Nevertheless, a clinician reading the article and focusing on the discussion section would note the following: “Based on the result that there is the same efficiency of the B. dracunculifolia and already marketed mouthrinses, we suggest the use of this natural substrate for prevention and reduction of dental biofilm, as well as caries disease.…”. This statement in the discussion section of the article misleadingly suggests equivalence in terms of treatment effect between the mouthrinses but also refers to a reduction in the incidence of caries disease, an outcome that the investigators did not measure.

3. Beware of large treatment effects presented in trials with few events. Clinicians often are appropriately skeptical of using evidence from the results of only 1 study and applying it in clinical practice. One argument is that the first studies that investigators conduct to determine the effects of an intervention usually have a

small sample size (for example, fewer than 200 participants) and too few events. A meta-epidemiologic study published in the oral medicine and maxillofacial surgery literature showed that the investigators of small randomized trials (that is, those involving fewer than 200 patients) were more likely to report larger and more beneficial effects compared with the investigators of large randomized trials (that is, those involving at least 200 patients) (OR = 0.92; 95% CI, 0.87-0.98; P = .009). Most of the time, our therapeutic interventions target 1 or 2 of the many pathologic mechanisms involved in the genesis of a disease. This is why, not only in dentistry but also in medicine in general, few interventions are able to demonstrate a large and real treatment effect.

The results of a systematic survey whose investigators analyzed 85,000 meta-analysis results extracted from 3,082 systematic reviews showed that, in 10% of the cases, the results of the first trial showed statistical significance and a large treatment effect, which afterward proved to be much smaller in comparison with the results that the investigators initially reported. It is important to notice that, when few events are available, even systematic reviews, including meta-analyses, could have this problem. Readers applying this guideline should beware of treatment effects that look “too large to be real,” because they are likely to be misleading.

To avoid this error, we suggest that clinicians and researchers should focus on CIs and minimal important difference estimates rather than P values in their interpretation of results. CIs provide a range of values, within which the true treatment effect is likely to lie, given the data observed in that particular study. Therefore, using CIs can help clinicians move away from considering trial results to be merely positive, neutral, or negative.

**Example: Is laser therapy effective for reducing facial swelling after sinus lift surgery?**

Investigators conducted a randomized controlled trial to determine the effect of neodymium-doped yttrium aluminum garnet (Nd:YAG) laser used for low-level laser therapy (LLLT) on pain, oral health–related quality of life (OHRQoL), and swelling after sinus lift surgery. The authors of the study concluded that “the application of Nd:YAG laser for LLLT was significantly effective in reducing postoperative swelling.” In the abstract of the study, they also mentioned, “We observed that the swelling and the OHRQoL in the Nd:YAG group were significantly improved when compared with the control group on the third day after surgery (P < .05).” A clinician not familiar with the concept of patient importance may conclude that the laser therapy resulted in an important reduction in facial swelling. A clinician who knows to focus on the magnitude of effect would note a graph showing the results for facial swelling on the third day (expressed in millimeters) in which improvement in facial swelling was only a difference of 2 mm. This difference, although statistically significant, represents a negligible benefit from the patient’s point of view. This observation contrasts with the study authors’ claim that the Nd:YAG laser for LLLT was significantly effective in reducing swelling.

4. **Beware of statements of statistical significance that claim clinical significance.** For decades, researchers have been using P values (that is, hypothesis testing methods) to determine whether there is an association between a risk factor and an outcome or to determine whether an experimental intervention applied to 1 group produces better health outcomes than a control intervention. This P value, although it tells us whether chance may explain differences between interventions, provides no information about the magnitude of the effect or the importance of the findings. Thus, readers who interpret small P values as large treatment effects usually are making a mistake.

5. **Beware of differences that are not statistically significant being interpreted as equivalence.** A common piece of advice that clinicians hear when using evidence to inform clinical decisions is that the “absence of evidence is not evidence of absence.” By convention, P values less than .05 are considered statistically significant, whereas values greater than .05 are called “not significant.” A misguided interpretation of results that are not statistically significant (P > .05) is that the results of a study have demonstrated that there is no important difference between the interventions being tested. Failure to demonstrate a difference does not, however, mean that an important difference does not exist. The sample sizes used in randomized controlled trials often are inadequate, resulting in a lack of power to detect a real and important difference that may exist. The investigators of underpowered trials (that is, trials with a small sample size and a small number of events) often are destined to fail to find statistically significant differences when comparing 2 interventions. Even when differences fail to reach the conventional P value threshold of .05, clinicians should not necessarily conclude that interventions are equally effective. The conclusion that no important difference
exists usually requires a large sample size. The CI is the best test of whether a sample size is adequate; a CI with a wide range indicates that a reader cannot conclude that no important difference exists. If the upper and lower values of the CI are close together—and neither, if representing the true effect, would constitute an important difference—only then is the conclusion of no important difference warranted.37

**BOX 6**

**Example: Does the primary closure technique result in fewer postoperative bleeding events than the secondary closure technique after mandibular third-molar extraction?**

The investigators of a systematic review of randomized controlled trials addressed the impact of secondary versus primary closure techniques on the occurrence of the postoperative outcomes of pain, swelling, trismus, infectious complications, and postoperative bleeding.71 Of the more than 14 studies that met the eligibility requirements, the investigators of only 4 studies had provided partial information regarding the incidence of postoperative bleeding. This example illustrates the fact that researchers in this area have overemphasized the importance of the beneficial outcomes compared with adverse events such as bleeding. The published results of the identified trials did not allow the authors of the systematic review to provide high-quality estimates for postoperative bleeding that would have facilitated clinical decision making.

6. Beware of uneven emphasis on benefits and harms. Clinical decision making requires the simultaneous consideration of many elements: certainty in the evidence, patient values and preferences, resource utilization, and the balance between benefits and harms.39 The results of randomized controlled trials sometimes inform the effectiveness and safety outcomes in a way that patients and clinicians can see both desirable and undesirable consequences of an intervention. Unfortunately, the investigators of many clinical trials do not report or do not measure adverse effects.40 In other cases, investigators present data about adverse events poorly; for example, the investigators may not provide any event rates for the treatment and the control arm, or they may fail to report the specific definition of the adverse event outcomes. Clinicians using this guideline should actively look for adverse event outcomes that are relevant for decision making. When these outcomes are not available, clinicians should acknowledge this as a major limitation of the results of a study.

7. Beware of misleading subgroup analyses. The investigators of clinical studies usually report the average estimates for the specific group of participants under study. Clinicians, on the other hand, try to personalize their prescriptions and indications as much as possible. One way that the investigators of clinical studies may provide the closest evidence possible to the patient is by using subgroup analysis. This type of analysis, conducted in primary studies and systematic reviews, aims to identify a specific subgroup of patients from the population who may respond differently to the treatment than other groups.43 Although subgroup analyses can be helpful, this type of analysis also can seriously mislead clinicians.

One example of how misleading subgroup analysis can be is the results of the Second International Study of Infarct Survival.44 The investigators presented an apparent subgroup effect in which patients who had a myocardial infarction and who were born under the zodiac signs of Gemini or Libra experienced an increase in cardiovascular mortality from aspirin use, whereas patients born under other zodiac signs experienced a benefit from aspirin use. No clinician would credit such a subgroup effect; the authors used this example to illustrate the dangers of subgroup analysis.

For clinicians trying to determine to what extent he or she can trust the results of a subgroup analysis in primary studies, researchers have proposed that clinicians ask themselves the following 4 questions43:

- Is it possible that chance can explain the subgroup difference?
- Is the identified subgroup difference consistent across studies?
- Was the subgroup hypotheses among the few tested and specified a priori?
- Is there a solid pre-existent biological rationale to justify the subgroup difference?
In addition to these questions, clinicians also should consider the following question to determine the credibility of subgroup analysis in meta-analyses\(^1\): was the comparison between subgroups conducted within or between studies? For detailed explanations of these criteria, we suggest referring to guidance published elsewhere.\(^9\) For clinicians applying this guideline, the message is simple: when presented with a subgroup analysis difference, remain skeptical until the results of additional studies confirm the hypothesis.

**Box 7**

Example: Is scaling and root planing more effective for reducing preterm birth in high- versus moderate-risk group patients?

The authors of a systematic review summarizing the evidence on the effect of scaling and root planing (SRP) in reducing preterm birth and risk of low birth weight conducted a subgroup analysis to explore potential reasons to explain the identified heterogeneity.\(^1\) The review authors conducted a post hoc subgroup analysis and set an arbitrary threshold of 22% risk of prematurity of the populations to divide the studies into 2 groups: a high-risk group (relative risk [RR] = 0.66; 95% confidence interval [CI], 0.54-0.80; \(P < .0001\)) and a moderate-risk group (RR = 0.97; 95% CI, 0.75-1.24; \(P = .79\)). The authors concluded that there was a “…statistically significant effect in reducing risk of preterm birth for SRP in pregnant women with periodontitis for groups with high risks of preterm birth only.” Although the test for interaction showed statistically significant results (\(P = .02\)), the authors conducted the subgroup analysis in a post hoc manner (that is, the cutoff point chosen to create the subgroups was not justified, and the difference was established at the level of “between studies” instead of “within study”). The authors of the review appropriately mention in their conclusions that “future research should attempt to confirm these findings and further define groups in which risk reduction may be effective.” Clinicians using this guideline should remain skeptical regarding this potential subgroup effect shown in the review and wait for more compelling evidence of such a subgroup difference.

**Conclusion**

Although clinicians have available a number of guides to critically appraise the risk of bias associated with clinical studies, little guidance exists addressing how to protect clinicians from being misled by the interpretations offered by the authors of clinical studies. In this final article in the EBD series, we present 7 criteria that clinicians can apply to avoid perpetuating misguided interpretations of study results. Clinicians should use these criteria to complement the guides provided in the other articles previously published in this series.\(^1\)

Dr. Carrasco-Labra is an instructor, Evidence-Based Dentistry Unit, Faculty of Dentistry, University of Chile, Santiago, Chile, and a doctoral student, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada. Address correspondence to Dr. Carrasco-Labra at Sergio Livingstone Pohlhammer 943, Santiago, Región Metropolitana, Chile, e-mail alonsocarrascolabra@gmail.com.

Dr. Brignardello-Petersen is a lecturer, Evidence-Based Dentistry Unit, Faculty of Dentistry, University of Chile, Santiago, Chile, and a doctoral student, Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada.

Dr. Azarpazhooh is an assistant professor, Dental Public Health, Faculty of Dentistry; an assistant professor, Endodontics, Faculty of Dentistry; and an assistant professor, Clinical Epidemiology and Health Care Research, Institute of Health Policy, Management and Evaluation, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada. He also is the head, Endodontics, Mount Sinai Hospital, Toronto, Ontario, Canada.

Dr. Glick is a professor and William M. Feagans chair, School of Dental Medicine, University at Buffalo, The State University of New York, Buffalo, NY. He also is the editor of The Journal of the American Dental Association.

Dr. Guyatt is a distinguished professor, Department of Clinical Epidemiology and Biostatistics, and a joint member, Department of Medicine, McMaster University, Hamilton, Ontario, Canada.

**Disclosure.** None of the authors reported any disclosures.


