


REVIEW

The OHStat Guidelines for Reporting Observational Studies and Clinical Trials in Oral Health Research: Explanation and Elaboration

A.M. Best¹ , T.A. Lang², B.L. Greenberg³, J.C. Gunsolley⁴, E. Ioannidou⁵, and Task Force on Design and Analysis in Oral Health Research*

Abstract: Adequate and transparent reporting is necessary for critically appraising research. Yet, evidence suggests that the design, conduct, analysis, interpretation, and reporting of oral health research could be greatly improved. Accordingly, the Task Force on Design and Analysis in Oral Health Research—statisticians and trialists from academia and industry—empaneled a group of authors to develop methodological and statistical reporting guidelines identifying the minimum information needed to document and evaluate observational studies and clinical trials in oral health: the OHStat Guidelines. Drafts were circulated to the editors of 85 oral health journals and to Task Force members and sponsors and discussed at a December 2020 workshop attended by 49 researchers. The final version was subsequently approved

by the Task Force in September 2021, submitted for journal review in 2022, and revised in 2023. The checklist consists of 48 guidelines: 5 for introductory information, 17 for methods, 13 for statistical analysis, 6 for results, and 7 for interpretation; 7 are specific to clinical trials. Each of these guidelines identifies relevant information, explains its importance, and often describes best practices. The checklist was published in multiple journals. The article was published simultaneously in JDR Clinical and Translational Research, the Journal of the American Dental Association, and the Journal of Oral and Maxillofacial Surgery. Completed checklists should accompany manuscripts submitted for publication to these and other oral health journals to help authors, journal editors, and reviewers verify that the manuscript provides the information

necessary to adequately document and evaluate the research.

Keywords: publishing/*standards, research design/standards, statistical data interpretation, comparative studies, retrospective studies

Introduction

“Large proportions of articles contain errors in the application, analysis, interpretation, or reporting of statistics or in the design or conduct of research” (Lang and Altman 2013). Oral health research is not immune to this criticism. For example, a 2009 review of 95 randomized controlled trials (RCTs) published in the leading journal in each of 6 dental specialties found generally suboptimal reporting of key Consolidated Standards for Reporting Trials (CONSORT) guidelines (Pandis et al. 2010). In another review, “spin”—

DOI: 10.1177/23800844241247029. ¹School of Dentistry and Department of Biostatistics, School of Medicine, Virginia Commonwealth University, Richmond, VA, USA; ²University of Chicago Medical Writing Program, Chicago, IL, USA; ³Epidemiology and Biostatistics, Touro College of Dental Medicine at New York Medical College, Valhalla, NY, USA; ⁴School of Dentistry, Virginia Commonwealth University, Richmond, VA, USA; ⁵UCSF School of Dentistry, San Francisco, CA, USA. Corresponding author: E. Ioannidou, UCSF School of Dentistry, 707 Parnassus Ave, San Francisco, CA 94143, USA. Email: effie.ioannidou@ucsf.edu

*Founded in 1968, the Task Force on Design and Analysis in Oral Health Research is a nonprofit, advisory organization of clinical researchers, basic scientists, biostatisticians, epidemiologists, and other quantitative scientists from universities, private research centers, government, and industry with experience in oral health research or clinical trials.



Article reuse guidelines: sagepub.com/journals-permissions

© Task Force on Design & Analysis in Oral Health Research 2024

nonstatistically significant results reported as “clinically important”—was assessed in the abstracts of 75 RCTs published in 10 leading dental journals. Of the 75 trials, 17 incorrectly presented a “statistically nonsignificant result for the primary outcome as showing treatment equivalence or comparable effectiveness” and 2 emphasized the conclusions of a secondary outcome when the primary outcome was not statistically significant (Roszhart et al. 2020). Additionally, a report of quality and spin in RCT abstracts in the periodontal-cardiovascular field found poor adherence to CONSORT guidelines, with 87% of trials not reporting on the primary outcome and 86% of trials showing at least 1 form of spin in the results and/or conclusions (Shaqman et al. 2020). Thus, “overall, dental journals show low reporting of quality-related characteristics with high variation that is journal-dependent” (Pandis et al. 2011).

Although oral health research is similar to clinical research in other fields, many dental studies have design characteristics that can confound analysis. For example, the unit of analysis can be a single tooth, multiple teeth, individual tooth sites, or a single patient. In longitudinal studies, teeth can be lost without disqualifying the participant from the study, and perhaps uniquely in human research, observational units may be added through the primary and permanent dentition process. Another unusual study design in oral health research is the split-mouth study (Lesaffre et al. 2009). A review of 119 such studies found improved reporting across 2 decades, but overall quality “was still below the acceptable level”: 85% did not provide a sample size calculation, 76% did not identify a primary outcome, 61% used inappropriate statistical methods that did not consider the correlated data, and 38% did not justify the design (Qin et al. 2020).

A common approach to improving reports of biomedical research is to use a checklist of reporting guidelines. Checklists can remind authors to report key elements of a study and help reviewers find where each guideline is

addressed when evaluating a manuscript. Most such guidelines are modeled after the CONSORT Statement for reporting randomized trials, first published in 1996 (Begg et al. 1996) and most recently updated in 2010 (Schulz et al. 2010). Also of interest to this document is the STROBE Statement for reporting observational studies (von Elm et al. 2014). Use of the CONSORT Statement has been associated with improved reporting of RCTs (Moher et al. 2001; Plint et al. 2006). However, the EQUATOR Network website lists over 550 checklists (University of Oxford Center for Statistics in Medicine n.d.). Thus, there appeared to be a need for a consolidated guideline that could address the main issues in the most common study designs in oral health.

Accordingly, members of the Task Force on Design and Analysis in Oral Health Research (Task Force on Design n.d.) began to develop guidelines for reporting clinical studies in oral health in 2019. The process of development is described in the OHStat Statement (Best et al. 2024). Drafts were circulated to editors of 85 oral health journals and to Task Force members and sponsors. The draft was discussed at a December 2020 workshop, attended by 49 researchers. The revision was circulated to the writing group and approved by the Task Force. As with other guidelines, the recommendations for reporting oral health research should 1) inform authors of the information needed to document and publish their research, 2) allow readers to assess the validity of the research or at least the credibility of the authors, 3) make the research process transparent, and 4) ideally, provide links to the information needed to replicate the study.

The target audiences for the OHStat Guidelines are authors, reviewers, and journal editors. Authors are advised to include the completed OHStat checklist when submitting a manuscript for publication. Journal editors and reviewers may also wish to consult these and other guidelines when evaluating a manuscript and should insist on complete adherence to the guidelines

within journal page limits, word limits, or in supplemental information. Critical appraisal and interpretation of observational studies and clinical trials in oral health will improve with an understanding of the details that support study validity. The purpose of this article is to provide the rationale and scientific background for each item. The terminology used is that provided in the original CONSORT Explanation and Elaboration document (Altman et al. 2001).

The OHStat Statement: Explanations and Elaborations

Identifying Information

The primary purpose of identifying information—the title and abstract—is to help readers make an informed choice about whether to read an article. Not so obvious is that this information should also help readers decide not to read an article. Thus, titles should identify the relationship that was studied. The title should not attempt to “capture the reader’s attention” with anything other than an accurate description of the research. Abstracts should not “highlight the research” but, again, should summarize it accurately so readers will know what to expect if they read the article (Lang 2010).

1. *Title: Space permitting, identify the research design in the title.*

The strength of evidence for health care interventions is limited by the study design. Including this information in the title helps with critical appraisal by assisting readers decide whether to read the article. Character limits notwithstanding, try to include as many of the SPICED-T elements as possible: Setting, Patients, Intervention, Comparator, Endpoint, Design, and sometimes Time frame (Lang 2020). A title can easily be shortened by removing the least important element. If applicable, some key elements must always be included in the title and abstract (e.g., single-sex studies).

2. *Abstract: Provide a structured abstract, as specified by the journal.*

The International Committee of Medical Journal Editors (ICMJE) recommends including a structured abstract when reporting original research (International Committee of Medical Journal Editors 2018). Such abstracts have 5 or more headings, and journals may specify which headings to use. Usually, only the results and conclusions require complete sentences. However, the form of the abstract will be specified by the individual journal.

3. *Consistency: Confirm that all information in the abstract is identical to that in the article, especially the conclusions.*

Many studies have found important discrepancies between the abstract and the full article (Lang 2022). Because abstracts are often separated from the full article, the information they contain needs to be identical to that in the full article. The conclusions, results, and objectives all need to be consistent throughout the manuscript.

The classic IMRaD structure of scientific articles (Introduction, Methods, Results, and Discussion) is well known, and the OHStat Guidelines emphasize the reasons for this organization. In 1965, Sir Austin Bradford Hill stated in an editorial board meeting of the *BMJ* that the structure of a scientific paper is built around the answers to 4 questions: “Why did you start, what did you do, what did you find, and what does it mean?” (Hill 1965).

Introduction: Why Did You Start? (Hill 1965)

After the title, the introduction is the most important and least-appreciated part of the scientific article. A good introduction can be enormously useful because it prepares readers to understand the paper, orients them to the research by establishing the need and importance of the study, indicates in general how the need was addressed,

and tells readers what to expect if they continue to read the article.

4. *Problem: Describe the background, nature, scope, and importance of the problem addressed by the research.*

Describe the historical, social, medical, ideological, or public health contexts of the problem. Indicate how serious and prevalent it is, as well as its consequences, implications, and whom it affects.

“Little is known about . . .” is rarely a good justification for doing research. A simple lack of knowledge is not sufficient to explain why a relationship needs to be studied or why a research report should be taken seriously (Lang 2017). Novice readers may need the background to understand the problem; experts expect a compelling justification of the research. The background in the Introduction should support a problem statement—the gap in knowledge or an untapped potential—that stimulated the research.

5. *Objectives: State the specific research objectives, including any prespecified hypotheses, in terms of a clinically important outcome measure or measures.*

The problem statement in the Introduction should support the choice of the primary outcome—the variable whose change in value is of interest and why it is clinically or practically important. The specific and measurable objectives should determine the methods of the research.

Methods: What Did You Do?

The purpose of the Methods section is to tell how the research question was addressed. The thought that a clear and transparent Methods section would allow someone to replicate the study is laudable but often not realistic, given the word limitations of a typical journal article, even with supplemental information. Instead, it may be better to tell readers where to obtain copies of the

protocol, the statistical analysis plan, and the original data set. In an article, a more reasonable goal is to provide enough information to establish the adequacy of the methods and, in so doing, establish the credibility of the authors as careful and thoughtful researchers.

6. *Design: Describe the overall study design and any variant (e.g., split-mouth, crossover, equivalence) and planned subgroup analyses.*

To understand the essential aspects of the study, its design should be described in the Methods. The hierarchy of evidence for clinical studies (both observational studies and clinical trials) arranges sources of information and research designs from those with the most control over error, confounding, and bias to those with the least control. We encourage researchers to aim for the highest appropriate level of evidence (American Dental Association [ADA] 2013; Oxford for Evidence-Based Medicine Group 2013). The hierarchy listed below is one of many versions, although all include essentially the same designs in the same order (Torabinejad and Bahjri 2005):

- Meta-analysis of RCTs
- Systematic reviews
- RCTs
- Cohort studies
- Case-control studies
- Cross-sectional studies
- Case series
- Case reports

At a minimum, authors should report whether an observational study is a cohort, case-control, or cross-sectional design and include information about the study timeline and variants such as nested designs or crossover studies.

The hierarchy of evidence should not be confused with the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) “system of rating quality of evidence and grading strength of recommendations” (Guyatt et al. 2011). The hierarchy simply

ranks therapeutic study designs by their potential to control for bias. Grades or levels of evidence usually refer to ways to describe or score the quality of individual studies.

A clinical trial is a “a research study in which one or more human subjects are prospectively assigned to one or more interventions (which may include placebo or other control) to evaluate the effects of those interventions on health-related biomedical or behavioral outcomes” (U.S. National Institutes of Health 2014). A clinical trial is one that meets all 4 of the following criteria:

- It involves human participants.
- It prospectively assigns participants to an intervention (not necessarily random assignment).
- It evaluates the effect of the intervention on the participants.
- It has a health-related biomedical or behavioral outcome.

Note that the definition includes both conventional parallel-group studies—where participants are assigned to interventions—and within-person studies—studies in which specific body parts or locations, such as lesions or dentition in the same individual, are assigned to experimental groups. The latter type of assignment allows participants to receive 2 or more treatments on different structures or areas so that each patient acts concurrently or sequentially as their own controls. Examples include dental split-mouth studies.

Conventionally, there are 2 types of human studies with health-related outcomes—clinical trials and observational studies. Most studies in oral health are observational studies, studies that do not meet the above definition.

The 7 guidelines especially relevant to clinical trials are in bold: 7, 8, 18, 19, 26, 27, and 39.

- 7. Approach: In a therapeutic clinical trial, say whether the study was intended to assess the intervention under ideal**

and controlled circumstances (an explanatory trial assessing efficacy) or under real-world conditions (a pragmatic trial assessing effectiveness).

Although the 2 designs have much in common, they differ greatly in terms of how they are designed and how their results are evaluated. An explanatory RCT evaluates the efficacy of an intervention under controlled conditions in a narrowly defined patient population, which maximizes internal validity but can limit generalizability (external validity). A pragmatic RCT is conducted under real-world conditions, where key aspects of the study can have great variability: the diagnosis, enrollment, treatment, participant adherence to treatment, and data collection. A pragmatic trial is designed to assess comparative effectiveness in more typical settings—to maximize external validity (Chalkidou et al. 2012). Other approaches (e.g., safety trials, dose finding studies) should also be reported, where appropriate.

- 8. Registration: If the study is registered, name the registry and give the registration number. State whether the trial was registered before the first patient was enrolled and whether the statistical analysis plan was determined before the data were analyzed.**

Both clinical trials and observational studies may be registered, thus improving transparency. Most major journals require, as a condition of publication, that a randomized clinical trial must be registered on a public trials registry, such as clinicaltrials.gov or others listed on the WHO International Clinical Trials Registry Platform (ICTRP) (World Health Organization 2014) before the first patient is enrolled (International Committee of Medical Journal Editors 2018). The basics of a statistical analysis plan (Gamble et al. 2017) should be outlined in the trial registration. Implemented originally in response to the suppression of negative studies, trial registration also allows the research

design and activities to be complete, detailed, and transparent and, hence, replicable. Registration of the planned study also permits comparison to the published study.

Observational studies also benefit from a statistical analysis plan. Outlining the details of study methodology before the study begins can distinguish a priori comparisons from post hoc analyses. Observational studies should clearly specify the hypotheses intended to be tested and the statistical methods planned to test them. Differences between the study as planned and the study as published should also be disclosed, as they should be in any design, and the differences and their effects on the reliability of the study results explained.

If applicable, tell how to obtain the study protocol, the statistical analysis plan, the original data, or any biological samples.

- 9. Ethics: Name the institutional review board that approved the study and give the study identification number. If the study was exempt from review, so state. State whether written informed consent was obtained from participants. Identify any competing interests of the authors and their employers.*

A standard requirement for conducting and publishing any research involving human and animal participants is prestudy approval by a recognized institutional review board (IRB), whose task is to protect the rights and welfare of participants during and after the study.

Some types of research, such as surveys, benign behavioral interventions, and routinely collected clinical or educational data, might be exempt from IRB approval, but such an exemption still needs to be approved by an IRB and documented in the article.

Prospective studies of adults must obtain written informed consent; studies of minors may be required to obtain assent. If relevant, describe the conditions under which consent was obtained. If the process of obtaining consent might be seen as intimidating

or coercive, describe the circumstances and the implications for the study. Compensation for participation must be disclosed.

Authors (and the authors' institution or employers) should report any competing or potential conflicts of interest that might influence or bias the conduct or reporting of the research (World Association of Medical Editors n.d.). Competing interests may be related to financial, professional, intellectual, political, or personal relationships. They may be only potential or perceived, or they may be factual. Competing interests do not necessarily mean that the research is biased. This information may be placed at the beginning or end of the Methods section or before the references, depending on the journal.

10. Funding: Indicate who funded the study and any role the funder had in planning the study, providing products or technical support during the study, analyzing the data, or publishing the results. Identify any competing interests of the funders.

Several groups fund clinical research, including government agencies, consumer groups, advocacy groups, private foundations, wealthy individuals, clinical centers, and industry. Almost any funding source has competing interests, that is, economic, programmatic, or reputational incentives to report favorable results. In addition, favorable results may also affect the chances of continued funding. Thus, the involvement of a funder in any phase of the research should be disclosed; any supplies, drugs, equipment, technical support, or unrestricted funding provided should be acknowledged. If no support was given, a simple statement to that effect is sufficient.

If participating individual physicians, group practices, clinics, or research sites were compensated for their time or contribution to the research, this must be disclosed.

Importantly, the potential for bias does not necessarily mean that the results are biased.

11. Setting: Indicate the setting(s) and location(s) of the study.

The setting or venue of a study (e.g., private practice, community hospital, academic medical center) and its location (e.g., rural, inner city, geographic region, country) can affect how well its findings might apply to other settings and locations. Aspects of location include social, cultural, economic, political, and geographic factors. If relevant to the generalizability of the findings, state why the setting(s) and location(s) were chosen.

12. Eligibility: Describe the population of interest. Give the criteria for eligibility.

The challenge of generalizability arises because your study includes past information on individual people, experiencing interventions, measured uniquely, at your particular location. These “instances on which data are collected” are only of scientific interest to the extent that your study results may generalize to the units, treatments, variables, and settings not directly observed (Shadish et al. 2002). Provide sufficiently detailed information to inform readers who was eligible for the study and to assess to whom the findings can be generalized.

For example, as recently as 30 y ago, women of childbearing age were excluded from clinical trials by Food and Drug Administration (FDA) policy. Current federal policies encourage including both sexes and all gender identities in clinical studies (De Castro et al. 2016; Wainer et al. 2020). Data should be routinely disaggregated and analyzed by sex or gender, as appropriate (Heidari et al. 2016). Single-sex studies must be justified if the reasons are not obvious. Women and minority group members are markedly less likely to volunteer for some clinical trials, so additional recruitment efforts may be required to achieve generalizability, inclusivity, and equity (Masood et al. 2019).

If race, ethnicity, sex, primary language, or disability is reported, indicate the classification options used and report

who assigned the categories (e.g., self-report, investigator judgment) (Dorsey and Graham 2011).

13. Recruitment: Tell how participants were recruited or identified. If done, describe any stratification or matching.

As item 11 identified *where* participant recruitment occurred, this item identifies *how*. Report the methods for participant selection/identification so that practitioners can decide how well their patients match those included in your study. Additionally, in combination with item 12 (eligibility), readers can also assess how well the study participants match the population of interest. The methods for case ascertainment and control selection are critical for evaluating case-control studies.

Typically, stratification or matching is employed during the recruitment process to ensure comparability of study groups. If done, give the reader sufficient information to judge the adequacy of these efforts.

14. Interventions: Describe the interventions or experimental conditions—including control conditions—and the protocol under which they were delivered.

Such descriptions might include the pharmacological properties of a drug, the technical aspects of a procedure, or patient home-care instructions. The most common types of comparators include placebo, a competing intervention, usual or standard of care, and untreated or unexposed. If multiple interventions occurred, describe the sequencing. Each intervention must be completely and accurately described if the research is to be evaluable.

This item describes what interventions were done; the next item describes how the database records this and other features.

15. Variables: Clearly identify the primary outcome variable (the primary

response), important secondary outcomes, and explanatory variables (exposures, risk factors, interventions, confounders). State the duration of follow-up, if any.

The primary outcome drives the study's sample size calculation and statistical power, so it must be clearly specified and defined (International Committee of Medical Journal Editors 2018). If possible, use common definitions and established outcome measures, to make comparing results across similar studies easier. Secondary questions and outcomes may be posed, but because trials are generally not designed or adequately powered to address secondary or exploratory outcomes, authors should interpret the results carefully (Pihlstrom and Barnett 2010). Secondary outcomes should be labeled as exploratory unless they are clearly outlined in a prespecified analysis plan.

Although outcomes with clinical and practical relevance are preferred, surrogate outcomes may also be used. If so, the biochemical mechanism or epidemiologic rationale for their use should be clear. Describe the established relationship between the surrogate measure and the clinical endpoint, where possible.

If a composite outcome is used—where 2 or more variables are combined into a single outcome—results should be reported for each of its components, in addition to the composite variable. Consider whether the components are of similar importance. For example, are the counts of decayed, filled, and missing teeth comparable (Casamassimo et al. 2009)?

In studies measuring outcomes at various time points, specify the follow-up duration of primary interest. State how the comparison time was determined and whether comparisons were made at a prespecified time point.

16. Unit of observation: Name the unit of observation or analysis (e.g., tooth, region of mouth, patient). Justify the use of partial-mouth studies.

Participants may be considered independent for the purposes of statistical analysis, but sites measured within a

patient's mouth are dependent, meaning that the value of one measurement is correlated with another. For example, periodontal measures in the same mouth are positively correlated (Imrey 1986; Fleiss et al. 1988; Imrey et al. 1994). The result is that analyzing tooth- or site-level data as if they were independent underestimates variability and overstates statistical significance (Fleiss et al. 1988). Therefore, analyses should account for correlated data (this approach is described in item 24). The statistical method for analyzing correlated data should be described in oral health publications (see item 30), where correlated measurements are the rule rather than the exception (Sterne et al. 1988; DeRouen 1990; McDonald and Pack 1990; DeRouen et al. 1991; Albandar and Goldstein 1992; Smith and Hadgu 1992).

Data from partial-mouth examinations can underestimate disease prevalence. Disease severity is overestimated if data are restricted to high-risk segments of the dentition (Eke et al. 2010). Accordingly, partial-mouth examinations should be justified, reported, and discussed.

17. Clinical importance: Where possible, but especially in clinical trials, report the minimum clinically important difference for the primary outcome.

The ultimate goal of medicine is to improve personal and population health. So, research should focus on clinically important and practically useful outcomes. The National Institutes of Health defines clinically meaningful outcomes as a measure of how a patient feels, functions, or survives (Biomarkers Definitions Working Group 2001). Facial aesthetics, tooth retention, and oral function are key to oral health. A practically useful outcome should be well defined, reliable, measurable, interpretable, and sensitive to the effects of an intervention (Fleming and Powers 2012).

For example, the ADA clinical practice guideline on the nonsurgical treatment of chronic periodontitis interpreted a mean difference in clinical attachment loss (CAL) between treatment and

control using a "clinical relevance scale" (Smiley et al. 2015). Even if statistically significant, a CAL difference of 0.2 mm was interpreted as "zero effect." A difference in the range of >0.2 to 0.4 mm was interpreted as a "small effect," and this was the minimum clinically important difference used in the practice guideline.

Small differences between large groups are often clinically meaningless. A "positive" finding is statistically and clinically defensible if all the values in a 95% confidence interval (CI) around the resulting effect size exceed the minimum clinically important difference.

18. Assignment: In randomized trials, tell how the random allocation schedule was created, concealed, and implemented. Tell how patients were assigned to groups.

As item 14 describes *what* the interventions were, item 18 describes *how* it was determined who got what.

Not all clinical trials use random assignment, nor is concealment always possible (Friedman et al. 2015). Tell how interventions were assigned or allocated. In parallel group studies, the schedule indicates the group to which the next enrolled participant will be assigned. In within-person studies (e.g., split-mouth studies), the schedule indicates the location or ordering of the interventions. If interventions were assigned at random, tell how this was accomplished (i.e., with the use of a validated statistical software program or a table of random numbers). The unit of randomization may not be the unit of measurement.

Report how (or whether) the allocation schedule was concealed from study personnel to prevent group assignment from being intentionally or unintentionally manipulated.

Implementation refers to how a participant is assigned to a group without anyone knowing whether it is to the intervention group or the control group. Tell who generated the allocation schedule, who enrolled participants in the trial, and who assigned patients to groups (Schulz et al. 2010). Studies

with inadequate or unstated allocation concealment tend to have significantly better outcomes than those with appropriate allocation concealment (Schulz et al. 1995).

Allocation concealment keeps group assignment hidden until after patient recruitment; blinding can also keep assignments hidden during the intervention and after.

19. Blinding: In clinical trials, indicate who was blinded to what information and how blinding was implemented. If applicable, indicate whether the control intervention could be distinguished from the experimental intervention.

Certain forms of bias may be prevented by using blinding (e.g., selection bias, ascertainment bias, and expectation bias). In clinical trials, blinding is not required, but when it is not used, this should be clear. If any, report the methods to mask the interventions from study administrators, from participants, and from those measuring outcomes. Contrary to popular belief, there are no widely agreed-on definitions for which groups are masked in a “single-blind” or “double-blind” study, so these terms should not be used (Lang and Stroup 2020). Consequently, specify who was blinded to interventions. Examples include participants, care providers, and those assessing outcomes.

Describe the similarities and differences between a placebo or sham procedure and the active drug or the trial procedure. Testing the effectiveness of blinding after the trial is over is uninformative because the results cannot be separated from pretrial expectations of the success of the intervention (Sackett 2004). Instead, indicate whether the interventions could be distinguished by the participants or those assessing outcomes. Report how blinding was maintained and whether or how it may have been compromised.

In all blinded studies using clinical examiners, specify what the assessor was blind to. For example, in a periodontal therapy trial, it is preferable for the

assessment of end-of-study pocket depth to be done blinded to the baseline value, as well as to group membership. Report whether laboratory values (e.g., IL-6) were assayed blind to group membership.

20. Data collection: Tell how data were collected throughout the study. If patients or information were excluded during the study, describe how the exclusions were identified and the reasons for exclusion.

The process of capturing data—the operational details of turning a concept into an entry in a database—bears directly on the validity of the data collected. How this process occurs can improve (or limit) the completeness and accuracy of the information used for analysis. Report information sufficient for a reader to judge these important details and to reproduce the process in future research.

Survey instruments (regardless if they are conducted on paper, via phone, or electronically) should be identified or provided in supplemental material. Cite a reference for any validation studies or established rating scales used, and disclose any modifications. Report how subscales or dimensions were scored and indicate any important thresholds (e.g., an established “normal” range, “high” or “low” scores). If the scale uses a “total score,” consider the effect of missing values (specifically, a missing value should not necessarily be scored as zero).

Large databases—clinical, administrative, billing—are increasingly available for analysis but have several characteristics that must be addressed (Katz 1997). Information recorded for another purpose must be converted into a research database. Report how the original information was collected and how entries are used or combined into variables for the study. Specifically, describe the classification methods for interventions, exposures, outcomes, and confounders. Consider the risk of misclassification because medical records are limited in studying clinical topics

and often contain errors and omissions in clinically important areas (Hornberger and Wrono 1997).

For example, the lack of a CDT code (Code on Dental Procedures and Nomenclature) for caries in a database from a periodontal practice does not necessarily indicate noncarious dentition. That is, a missing value may lead to misclassification bias or to unmeasured confounding. Attend to time-stamped records to ensure that the values of predictor variables precede the encoding of outcomes (and not the other way around).

The codes and algorithms for subject selection should be either given in detail or made available on request, including how the algorithms were validated. Report how records are linked between databases. If the algorithms are extensive, consider including this information in supplemental material accompanying an article.

21. Measurement: Describe any steps taken to improve the quality and accuracy of measurements. For judgments, describe the assessors’ qualifications, as well as what they knew about the participant before making their judgment, and report the degree of agreement for their judgments.

Science depends on measurement (“Can’t measure it, can’t do science on it.”). Accordingly, describe any training, experience, calibration, monitoring, or other efforts to improve the accuracy and consistency of measurements. Indicate the number of measurements taken for each outcome of interest, the number of independent observers, and level of inter- and intraobserver variability (e.g., κ , percent agreement, intraclass correlation coefficients). Disagreements between assessors are common (Holtfreter et al. 2012).

Describe what assessors knew about the study participant before they rendered a judgment. Report whether clinical outcomes were determined by the same individuals implementing the intervention(s). The independent

assessment of predictors and outcomes adds to the credibility of findings.

22. *Threats to validity: Describe any procedures used to minimize error, confounding, and bias.*

The Methods section should identify potential sources of error, confounding, and bias and tell how these issues were addressed in the design or analysis. Describe how the role of potential confounders was addressed, including the use of stratification or statistical adjustment.

There are a large number of potential sources of bias to consider in the design, execution, analysis, and interpretation of research (Hartman et al. 2002). If your overview of the problem (item 4) identified potential bias in previous research, report your methods to overcome these difficulties.

Statistical Methods

23. *Sample size: Explain how the sample size was determined; specify the minimum clinically important difference in the primary outcome (effect size) and other values used in a power calculation.*

In hypothesis-driven research, but especially in RCTs, a sample size should be reported and based on an a priori power calculation. Report the assumptions made in the determination (e.g., effect size, estimates of variability, expected dropout rates). Where possible, include an estimate of the minimum clinically importance difference on the primary outcome. Calculations should consider confounding variables as well the implications of insufficient enrollment, dropouts, or missing data (Hsieh 1989). Split-mouth studies require additional documentation (e.g., the standard deviation of the within-person differences) for sample size calculations (Pandis 2012).

For example, the minimum clinically important difference of 9 points (out of 80) on an oral health quality of life scale guided the sample size determination in a removable partial denture framework study (Ali et al. 2020).

Power calculations to determine sample size are not required—for example, the analysis data may have been previously collected for another purpose. In studies where the sample is fixed, describe how the study size is sufficient to estimate clinically meaningful differences. If a study is “too small,” the confidence intervals may be too wide to make a meaningful conclusion; if a study is “too big,” clinically inconsequential differences may be found (Altman and Bland 1995).

Studies designed to test equivalence or noninferiority (or “just as good as” studies) differ from superiority trials (studies of differences). Among other things, equivalence studies require a prespecified range of clinically important therapeutic effects (i.e., the equivalence margins) that must be reported and justified. See the article by Piaggio et al. (2012) for sample size calculation in equivalence studies.

24. *Analytic approach: Identify the key statistical methods used to analyze the data.*

Statistical analyses should include a predetermined plan for analyzing the primary outcome, with specific objectives and clear plans for addressing secondary and exploratory aims. Specify the statistical software used (e.g., SAS). If necessary for clarity, briefly note the procedures/extensions used (e.g., PROC LOGISTIC). As needed, report details in supplemental information.

The goal is to describe statistical methods with enough detail to enable a knowledgeable reader to assess the validity of the results and for those with access to the original data to verify the reported results. The data set and computer code used to perform the analysis should be available if requested.

25. *Primary analysis: Explain how differences or changes in the primary outcome were analyzed; how associations were estimated.*

An RCT of a single outcome may rely on randomization to justify a simple

comparison reflecting the primary aim. An RCT with a longitudinal measure (e.g., baseline and follow-up) may also rely on randomization to ensure baseline comparability. By definition, only a longitudinal study may assess “change.” A repeated-measures mixed-model approach to account for baseline imbalance should be considered. In observational studies, analysis of covariance (ANCOVA) or adjustment for baseline values is generally inappropriate (Blance et al. 2007; Etminan et al. 2021).

Always report absolute risks because all other expressions of risk can be derived from these. Be aware that analyzing percent change can easily be misleading. Analyzing percent change (the difference from baseline divided by the baseline value) may violate several statistical assumptions and can easily exaggerate effects, so such analyses should be avoided.

For information on analyzing and reporting equivalence of noninferiority studies, see the articles by Piaggio et al. (2012), Flight and Julious (2016), and Ebbutt and Frith (1998).

26. ***Analysis populations: In randomized trials, indicate whether the analysis was by intention to treat, per protocol, as treated, or some combination. Describe exactly who was included in each analysis.***

Even in observational studies, it must be clear who was included in every analysis. Missing outcome data can be problematic but can be accommodated by modern statistical methods. In clinical trials, some patients may drop out, not receive the intended treatment, or not adhere to the trial protocol. To preserve the benefits of an RCT, intention-to-treat analysis (ITT) is recommended (Wood et al. 2004). Simply stated, ITT analysis means “once randomized, there analyzed” regardless of whether subjects actually received the allocated interventions or whether they adhered to follow-up visits or trial protocol. This analysis requires 2 conditions: all

randomized patients should be included in the analysis, and they should be analyzed in the group to which they were allocated.

For superiority trials, report the ITT analysis for the primary outcome (Lachin 2000). Exclusion of eligible participants for any reason is incompatible with the intention-to-treat principle and may bias the results. Accordingly, include all randomized participants in the primary outcome analysis. This conservative approach acknowledges that participants may drop out *because* of the protocol.

There is no consensus on acceptable “modified ITT” criteria (Brody 2016). Modified ITT is often reported inconsistently and has increased (Abraha and Montedori 2010). Deviations from ITT described as “modified ITT” may exclude patients who did not commence their randomized intervention (as-treated analysis), patients without a baseline assessment, patients without a postbaseline assessment, patients not returning for follow-up assessments, or patients found to lack a specific diagnosis at entry. Report the justification for modifications to the standard criteria. A per-protocol analysis includes only participants who completed the study without major departures from the protocol. Such analysis may be reported—it can indicate effectiveness—but should not supplant the ITT analysis (Shrier et al. 2014). Published reports of clinical trials should clearly distinguish between ITT analyses and all other forms by describing who was included in each analysis.

Many clinical studies analyze the data for 1 or more subgroups. Planned and well-specified subgroup analysis has a stronger basis for inference. In contrast, post hoc subgroup analysis is at high risk for spurious findings and is typically discouraged; at a minimum, it should always be identified as exploratory (Pocock et al. 1987; Mills 1993). For case-control and cohort studies, analyzing subsets of the study population that were not part of the original study objectives is not appropriate.

27. *Stopping rules: In clinical trials, describe any interim analyses or stopping rules and indicate who could stop the trial.*

The strongest inferences are made in trials that are completed as planned (i.e., reaching 1 or more of the following planned goals: obtaining an adequate sample size, collecting follow-up data from a sufficient number of patients, having event counts sufficient for analysis, or closing the trial on the scheduled date).

However, sometimes trials are stopped early when an interim analysis triggers a statistical stopping rule. Performing multiple statistical analyses as the data accumulate during a trial (usually for safety reasons) weakens inference and increases the chances of reporting spurious results unless appropriate statistical corrections are made. The timing of all interim analyses should be reported, as should adjustments made to account for multiple analyses (e.g., multiple comparison or group sequential methods) if the interim results are to be published.

As above, when an interim analysis finds that a treatment is exceptionally effective or exceptionally harmful, the trial may be stopped early. Withholding an effective intervention from the control group may be unethical, as is continuing to subject the treatment group to a harmful intervention. If applicable, report how an independent data monitoring committee examined the accumulating data and include any formal statistical stopping rules.

28. *Data preparation: Identify any data-cleaning procedures used to modify raw data before analysis (e.g., missing data, loss to follow-up, transformations, creating or combining categories, outliers). Clearly distinguish between prespecified modifications and those arising during analysis.*

In dental studies, missing teeth pose a unique problem. Before the study begins,

identify strategies to accommodate missing teeth (e.g., the measurement of a contralateral tooth). In longitudinal studies, when a tooth was measured at baseline but is no longer present at follow-up, it should not be considered “missing” in the statistical sense but rather could be considered a negative outcome. This problem is similar to that of what to do with patients who drop out of a clinical trial, except that in oral health studies, a tooth may drop out and the patient remain.

State specifically how missing data were handled in the analyses. Measures to prevent missing data and to retain participants should also be described. Missing data may be associated with loss of power and potential bias.

Unless missingness is rare, complete case analysis—excluding participants (or teeth) with missing data—is rarely justified. “There are no universally applicable methods for handling missing data” (Shih 2002). We recommend assessing the differences between comparison groups in retention rates and patterns of “missingness” and exploring characteristics likely to be associated with missing data or dropouts. Describe the analytic approach used to address missing data, including methods for imputation and any sensitivity analyses used to explore the potential impact of missing data.

Prematurely dividing a continuous distribution of values into 2 or more categories can reduce statistical power and may introduce bias, depending on how the categories are determined. Thus, continuous variables should be maintained as such during analysis unless well-established and accepted criteria justify categorization. Report why the categories were created, when they were created (before or after data collection), and where and how the boundaries were assigned. This guideline does not preclude categorizing variables after analysis, rescaling units to be more clinically meaningful or to simplify communication and promote clinical utility.

Details on data transformation and imputation should be included in the

statistical analysis section. If skewed data were mathematically transformed for analysis, indicate the transformation used (e.g., square root, log) and whether the transformation was successful (i.e., suitable for analysis with parametric methods). When describing the results, transform the results back to make them clinically meaningful (e.g., “square root follow-up time” should be back-transformed to months). If results are best expressed as percentage change, use the preferred method of analysis and then convert the summary statistics into absolute or relative risk (Vickers 2016).

29. *Multivariable modeling: Identify the purpose of analysis, the response and predictor variables considered, and the statistical procedures used in the model-building process.*

Describe the predetermined analyses plan for the primary outcome. List specific objectives and clearly address plans for secondary or exploratory aims.

The study design should drive the modeling approach of the specific aims. If variable selection is employed, it should follow a well-defined procedure to control for potential bias in the final model. If possible, determine whether interaction between predictors is present; if so, describe effect modification (Hyman 2006). For prediction modeling, all “candidate” predictors should be evaluated holistically (Steyerberg and Harrell 2016). If applicable, identify the variable-selection process used (Nguyen et al. 2019; Talbot and Massamba 2019). Be aware that data-driven methods have been shown to be biased toward too high an estimate with too narrow confidence limits.

30. *Correlated data: Tell how correlated data (e.g., nonindependent or paired) were treated in the analysis. More than one outcome measurement from the same participant (e.g., multiple teeth or across time) usually must be explicitly modeled in the analyses.*

In any study where there are multiple measurements on the same individual,

the correlation between these measures should be considered. In within-person trials, each participant is subjected to 2 or more treatments, and measurements are therefore correlated. In such trials, a group is the set of participants' body sites allocated to a particular intervention or to the order in which the interventions are given. Report the statistical methods appropriate for the specific within-person design employed (Pandis et al. 2019). Report the observed correlation between body sites for continuous outcomes and tabulation of paired results for binary outcomes. In these trials, the expected correlation of within-person treatment outcomes should be incorporated when estimating the sample size (Hujuel and Loesche 1990; Hujuel 1998). In designs in which segments, quadrants, or half-dentitions within each subject are assigned interventions, consider possible carryover effects (Chilton and Fleiss 1986; Hujuel and Moulton 1988; Lesaffre et al. 2009).

31. *Ancillary analyses: Describe any ancillary analyses (e.g., sensitivity analyses, data imputation, assessing assumptions of the analysis, interaction analysis, confounding).*

Ancillary analyses are intended to support the preplanned primary (and perhaps secondary) analyses. Analyses suggested by the data are addressed in item 32.

A sensitivity analysis can determine the robustness of the findings to changes in methods or assumptions. Subgroup or interaction analysis may be used to explore whether the findings are consistent in subpopulations. Missing data can lead to potentially biased results and loss of power. If missing values are imputed, document the prevalence of missing cases for each variable and describe the method of imputation. Multiple imputations require reporting the results of sensitivity analysis.

There is no consensus on how to assess the assumptions that underlie common analysis methods (Nørskov et al. 2021). Many assumptions cannot

be statistically established, and only context knowledge will serve to guide the analyses (e.g., see item 30, correlated data). The validity of results may be enhanced by reporting clear, complete, and transparent assessments—likely in supplemental material because of publication limitations.

32. *Post hoc analyses: Identify any post hoc or exploratory analyses, including unplanned subgroup analyses, and identify them as such.*

Whereas ancillary analyses may support the primary aims, analyses suggested post hoc by the data or initial analyses can only be considered exploratory. As Marcia McNutt, past editor of *Science* and then president of the National Academy of Sciences, said, we “have no problem with true exploratory science. . . . But it is important that scientists call it as such and not try to pass it off as something else” (Shell 2016).

If exploratory findings are to be reported (item 41), specify clearly the way the data were approached for the exploratory analyses. The limitations of post hoc analyses are explicit if the process is transparent.

33. *Hypothesis testing: If P values are reported, identify what is being compared, as well as the statistical test used for the comparison, and report the calculated P value (i.e., $P = 0.063$, not as $P > 0.05$ or NS).*

A credible scientific claim has many components, and statistical analyses continue to be critically important in supporting claims. However, researchers often rely on “ $P < 0.05$ ” as a bright-line indicator of success, which has led to overstated effects and associations and to understated uncertainty. Results, therefore, are often misinterpreted as being clinically meaningful because of a low P value (Gelman and Loken 2014). For this reason, several prominent statisticians recommend that the term *statistically significant* be abandoned (Wasserstein et al. 2019). Although

a spectrum of positions remains on this issue, we support the strong and longstanding consensus of the statistical community that the binary “yes or no” convention for denoting importance of a finding based on its P value alone is not logically supportable and should be discarded (Wasserstein and Lazar 2016).

That said, in carefully designed and executed clinical trials or in large population-based sampling studies, sometimes it is appropriate to base statistical inference on classical null hypothesis significance testing. In such cases, it is important to emphasize that a P value does not indicate probability, truthfulness, or importance (clinical or practical). For example, research data may indicate that mean periodontal clinical attachment level improved by 0.25 mm ($P < .0001$) or that the number of decayed, missing, or filled teeth worsened by 5 ($P > 0.2$). These P values do not mean that attachment level improvement is real, probable, or important or that 5 decayed teeth are not real, probable, or important. “No P value can reveal the plausibility, presence, truth, or importance of an association or effect” (Wasserstein et al. 2019). Statistical significance does not ensure scientific validity, and it does not indicate clinical importance (Best et al. 2016).

In most instances, the estimated result and its 95% confidence interval are preferred. A larger interval indicates a less precise estimate, so the range of the interval should also be considered in the interpretation of potential clinical importance. More important, an interval that contains both clinically important and unimportant values usually suggests ambiguous results and should be interpreted with caution.

The decision to accept or reject a manuscript based on “statistically significant” results should be replaced with criteria based on the strength of the study design and analyses. Similarly, outcomes having small P values should not be touted as being meaningful or important without considering other factors, such as the susceptibility of the study design and analytic methods to bias and the size

and importance of the outcome. After considering these and other factors that may explain an apparent association, a small P value indicates sufficient evidence to overcome chance as a plausible explanation of the result. On the other hand, a large P value simply indicates that there is not enough evidence to disregard chance as one explanation for the lack of an observed association.

Never report only “The results were statistically significant ($P < 0.05$).” In conventionally sized studies, in addition to a statement regarding the magnitude and direction of associations, report the calculated P value (Council of Science Editors 2015; Christiansen et al. 2020). Specifically:

- P values should usually be expressed to 2 or at most 3 decimal places. The smallest P value that needs to be reported in clinical medicine is $P < 0.001$.
- For large calculated P values, consider reporting the calculated value to 1 decimal place. For example, “We observed no evidence for a difference ($P > 0.9$).” Do not use the abbreviation “NS” (not significant).
- Discontinue designations for levels of significance—for example, a single asterisk for $P < 0.05$. Instead of using an asterisk, consider the confidence interval as a better description of effect size. If designations are used in a figure, the calculated P values should be available in text, table, or figure legend.
- With one exception, never report a P value as zero or 1. Report these rounded values as $P < 0.001$ and $P > 0.9$, respectively. Only report a P value as 1 if it is produced by an exact test (e.g., Fisher’s exact test) and is, in fact, exactly equal to 1.

Analyses of large databases often can produce very small probabilities ($P < 0.0001$) as artifacts of the large sample size. Accordingly, such studies should be interpreted by practical or clinically meaningful considerations (e.g., declaring clinical significance if the relative risk is greater than 1.2). Studies of genetic associations often include large numbers

of variables, and methods to control for false positives in these studies are not necessarily adequate. Currently, these methods include requiring P values to meet stringent thresholds to establish statistical significance (e.g., an α of 10^{-8} or less) or independent replication (Qu et al. 2010; Pulit et al. 2017).

Results: What Did You Find?

The obvious purpose of the Results section is to report and describe the findings of the study: the data that were collected and the relationships among them. A purpose just as important is to tell what happened during the study, such as protocol deviations, changes in the intervention, and unexpected data losses. Numbers in the text are difficult to read and compare, so data should be reported in tables or graphs whenever possible and duplicated in the text as little as possible (Council of Science Editors 2015; Christiansen et al. 2020). Ideally, call attention to general results in the text and refer readers to the details in tables and figures: for example, “Periodontal disease was present in 28% of the patients (Table 4)” (Lang 2010).

If there are changes in the protocols or other important revisions in the conduct of the study, describe them in the first paragraph of results. Typically, however, the Results section should begin with a description of the participants, include simple presentations of one-variable-at-a-time results, and end with results of a multivariable model. Sufficient detail should be provided so that results can be verified and integrated into other analyses (Lang and Altman 2013).

Although randomized controlled trials are considered the strongest research design because they provide the most control over bias, most studies are observational. Thus, the findings from observational studies should be phrased using terms such as “association” or “related.” Avoid terms that connote causality such as “lead to,” “effect,” “influence,” and “produce” unless the result arises from an appropriate analysis of a causal model (Bellamy et al. 2007).

Report the numerators and denominators for percentages, rates, and ratios. Summarize continuous data with a measure of central tendency and a measure of variability. For distributions that are reasonably symmetrical, means and standard deviations are appropriate. For nonsymmetric data, give the median and an appropriate percentile range. Do not use the standard error of the mean (SE or SEM) to describe the variability of observations. The SE is an inferential statistic, not a descriptive one. (A range encompassing an estimate ± 1 SE represents a 68% CI.)

34. *Participants: Report the number of participants included and excluded at each stage of sample selection, group assignment, at key times during the study (including those lost to follow-up), and the number analyzed in each group and subgroup (consider summarizing this information in a flow diagram).*

Visually summarize the research design and analysis populations in a flow diagram. The diagram can show the number of participants at each stage of sample selection, the size of each group and subgroup in the analysis, and the number of participants with various outcomes. Another organization for the flow diagram identifies a target population, a source population screened from the target population, an eligible sample selected from the source population, and the study participants enrolled from those eligible. Both the numerators and denominators for intention-to-treat and per-protocol analysis can be shown, for example. The flow diagram also allows all participants to be accounted for at each stage of the study by checking the numbers.

35. *Sample: Describe the sample; report baseline demographic and clinical characteristics, including measures of variability, for each group.*

Participant characteristics are often best reported in tables with standard

descriptive statistics. Generally, because items are more easily compared side-to-side (space permitting), groups should be named in the column headings, and the variables on which they are compared should be named in the row headings (Lang 2018). Column headings usually also indicate the size of each group (e.g., “ $n = 25$ ”), and row headings usually report the unit of measurement for the variable. Column and row totals can also be informative.

Report numbers and measurements with an appropriate degree of precision. Percentages are preferred to proportions. Numerators and denominators should always be clear and easily found. Round to whole numbers unless there is a compelling reason not to. Reporting more than 2 decimal places is rarely needed.

In an RCT, do not compare baseline differences with significance tests; by definition, any imbalances occurred by chance. Consider imbalance in light of the ability of the predictor to influence the outcomes. Consider incorporating clinically important imbalances into the analyses and report how the choice was made. In case-control and cohort studies, this is also important.

36. *Study periods: Define and give the inclusive dates or defining events of any distinct study periods (e.g., recruitment, data collection, outcome assessments, follow-up). Consider presenting this information in a timeline.*

Therapies and case definitions continuously evolve, and the exposures and risks in a community can change profoundly. Report the actual time frame of the study so that it may be compared to others.

37. *Results: Report the results of the outcome variables for each group; provide a measure of precision (95% CIs) for each comparison, focusing on the primary outcome. Distinguish within-group differences from between-group differences.*

Summarize the results in clinically meaningful or practical terms. For example, “Brushing with fluoride toothpaste had a statistically significant effect on the mean number of decayed, missing, and filled primary tooth surfaces (DMFS) . . . for populations at high risk of developing caries [standardized mean difference = -0.25 (95% CI = -0.36 to -0.14)]” reports the result as a standardized difference. Reporting in clinically meaningful terms would phrase it as “(DMFS difference = -1.92 (95% CI = -1.32 to -2.49)).” Include starting and ending values and a brief summary of an analysis for the primary outcome of interest, as well as any prespecified secondary outcomes identified in the methods and each of the primary covariates. All baseline and end-of-study descriptive statistics should be accompanied by appropriate measures of variability. As with group descriptions, group comparisons are usually best summarized in a table. As a result, the differences and their confidence intervals (and P values) will usually be reported in the right-hand column. In addition to reporting differences between groups on clinical outcomes, it is often useful to report the numbers or percentages of patients who did and did not improve.

There is a difference between reporting a “difference” between independent groups and reporting a “change” across time within a group—a difference reflected in the statistical analysis used. But there are any number of “within-group” studies in oral health; split-mouth and crossover designs come to mind. In these cases, it is important to report the repeated-measures analysis method employed and to report results accounting for the within-person correlation; see item 30: correlated data.

Tables and graphs used to collect or analyze data may not be optimal for communicating data. They should be designed to present patterns in the evidence, such as trends, differences, or associations, especially to clarify relationships that would otherwise be difficult to explain in the text. Tables can effectively summarize and compare

detailed information. Figures effectively show trends and patterns in the data (Council of Science Editors 2015; Christiansen et al. 2020).

Tables and graphs should complement rather than duplicate each other or the text. Because missing data are common, the sample size should be clear for every summary statistic. Consider including a column with the number of participants in a table. Tables and figures should be understandable without undue reference to text. Except for horizontal lines, tables should generally be free of lines, boxes, arrows, or other devices unless they indicate the structure of the data (Lang and Secic 2006; Lang 2010).

Clinical and laboratory images (e.g., radiographs, photographs, electrocardiograms, blots) differ from other visuals in scientific publications because they do not present, organize, or summarize information; they are the information. For this reason, images must be well documented. The 6 CLIP principles (Lang et al. 2012) identify key information that could or should be reported:

- Identify the subject of the image.
- Tell how the image was acquired.
- Explain why the specific image was selected.
- Describe any modifications of the image after it was obtained.
- Emphasize the important details of the image itself.
- Interpret and give the implications of the image.

The overarching goal is that an image must correctly and clearly represent the scientific content. However, the National Institutes of Health's Office of Research Integrity reports that more than 80% of accusations of misconduct involve image manipulation (Office of the Secretary 2017). Always retain the unprocessed image and clearly document all changes made to the submitted image. Follow journal guidelines for permissible processing. The most common problematic manipulations are undisclosed incidences of (Rossner and Yamada 2004)

- Splicing different images together into a single image
- Changing brightness and contrast on only part of the image
- Using cloning tools to hide details
- Cropping images to eliminate information

38. *Deviations: Report any changes in the protocol during the study.*

Describe participants who did not complete the protocol (e.g., those leaving the study, lost to follow-up, whose treatment was ended, and those who deviated from the protocol).

39. **Harms: In clinical trials, describe any adverse events or harms, including whether or not they might have been caused by the intervention.**

“Harm” is the totality of possible adverse consequences of an intervention or therapy; they are the direct opposite of benefits, against which they must be compared” (Ioannidis et al. 2004). Report expected and unexpected adverse consequences so that readers may make informed decisions about using interventions in practice. Even in observational studies, consider the effect of dropouts and loss to follow-up on the results. Adverse outcomes can impact the validity of the study or affect whether it is ethical to continue a longitudinal study.

Especially in RCTs, where harms may be caused by the intervention, report any harms or adverse events (Ioannidis et al. 2004). Describe or identify harms with standard definitions, including any grades for severity and extent, how they were detected, whether or not they were prespecified, whether they were anticipated or unexpected, and whether they were attributed to an intervention. Provide a balanced discussion of benefits and harms in the context of a study's limitations and generalizability. When necessary, report harms in supplementary tables.

40. *Modeling: Report the results of any multivariable modeling, including*

interaction terms. Consider how to best report the models in tables.

The results of unadjusted analyses may be reported, often as a prelude to the definitive findings from the adjusted—multivariable and/or multivariate—analysis. Unadjusted analyses should not be used for final interpretation unless confounding can be excluded.

For each multivariable analysis, report the measure of association or difference with corresponding confidence intervals for all variables in the final model, including any interaction terms. If the number of confounders or covariates is large, this detail could be included in supplemental material so that the summary table in the manuscript can focus on the primary factors of interest. Provide an appropriate measure of the model's goodness of fit to the data (e.g., R^2 for linear regression).

41. *Exploratory analyses: Report the results of any exploratory analyses (e.g., subgroups, interactions, sensitivity analyses) separate from the primary outcome results.*

Research results commonly suggest further analyses. These post hoc analyses must be interpreted more cautiously than planned comparisons. Subgroup analyses and comparisons especially must be interpreted carefully, given the reduced statistical power associated with smaller sample sizes and the increased number of hypotheses tests, which can create the multiple comparisons problem of false positives and can lead to claims of “data dredging” (Erasmus et al. 2022). A post hoc power calculation should not be reported as it provides no additional information (Christogiannis et al. 2022).

Discussion: What Does It Mean?

When writing the Discussion (and the cover letter to the journal), keep in mind the advice attributed to Franz Ingelfinger, editor of the *New England Journal of Medicine*: Answer the questions “So what?” (Is this research new, valid, important, and well reported?) and “Who

cares?" (Why do readers need to know about this research?).

42. *Summary: Summarize the study and the main results.*

Answer the research question posed in the Introduction. Briefly summarize the study but emphasize the final results. With a prespecified analytic plan, including adjustments for multiple comparisons, results may be reliably expressed for the primary analyses.

Avoid emphasizing results suggested by the data. Results from post hoc analyses should be labeled as descriptive or exploratory only and should be summarized separately.

Ensure that the main conclusions match those in the abstract (see item 3). Discrepancies in the information reported in the abstract with that reported in the article are distressingly common, serious, widespread, and longstanding (Zhang and Liu 2011; Bastian 2014; Lang 2022).

43. *Interpretation: Interpret the results cautiously and suggest an explanation for them. Separate the interpretation of the prespecified outcome analysis from post hoc analyses.*

Discuss both the expected and unexpected results. The estimated treatment effect should be accompanied by a measure of precision (typically a 95% CI) and should be interpreted in terms of clinical or practical importance (Brignardello-Petersen et al. 2013). The implications of both the lower and upper limits of CIs should be considered when assessing clinical or practical importance.

44. *Integration: Compare the results with what else is known about the problem; attempt to integrate the study with the literature.*

For each research question, compare and contrast the findings of others with the results presented in this study. Depending on the topic, references more than 5 or 10 y old are generally less

relevant, with the exception of seminal articles or comprehensive reviews. Cite the original source when possible; secondary sources are often incomplete and inaccurate. Read the full reference (not just the abstract) before citing it.

45. *Generalization: Discuss the generalizability of the results (their external validity).*

Describe the extent to which the study data may be representative of the population of interest (see item 12) (Shadish et al. 2002). In clinical trials, this may be affected by the approach (see item 7). Indicate how the results might be applied to other populations or settings. Generalizing often requires speculation, which should be acknowledged in the article.

46. *Implications: If reasonable, comment on the applications or implications of the results on health care delivery.*

If reasonable, speculate about how the findings might improve patient care if the intervention were to be widely adopted. If the results do generalize to other populations or settings, call attention to possible implications. For example, a more sensitive diagnostic test may detect more cases, increasing the number of patients treated and increasing the total treatment costs. A more effective but expensive treatment may not be affordable to the patients eligible to receive it. A new technology might require specialized maintenance capabilities and special training for those who use it.

Avoid saying that "more research is needed." More research is always needed. Instead, if possible, suggest specific ways in which future research might be improved.

47. *Limitations: Describe likely sources, direction, magnitude of error, confounding, and bias that were not controlled for in the study design or analysis. Do not cite the standard limitations of the study design.*

The Cochrane Collaboration has a useful tool for recognizing the main sources of potential bias: the ROBINS-I tool for assessing risk of bias in clinical studies (Sterne et al. 2016). Potential sources of bias include participant selection, unmeasured or uncontrolled confounding factors, inconsistent interventions, imprecise measurements, protocol deviations, missing data, variation in judgments, and selective reporting of results. The major limitations of retrospective and nonrandomized designs, self-reported surveys, analyses of databases clinical registries, and so on are widely known and need not be reported.

Many authors do not report limitations for fear their paper might be rejected. If limitations are acknowledged, a reviewer knows the authors were competent enough to recognize a limitation and honest enough to acknowledge it. Readers appreciate modesty as well.

48. *Conclusions: List the conclusions in terms of a clinically important outcome measure. Do not restate the results; give their implications.*

Listing each conclusion promotes specificity and helps readers better understand the research. Do not overstate the implications of the research and do not speculate on the conclusions.

Results are not conclusions. "We found a 65% reduction in dental caries" is a result, not a conclusion. A conclusion identifies the clinical or practical implications. For example, "We believe the data clearly support the use of this treatment in children at high risk for caries in supervised brushing environments."

Conclusions from clinical trials—randomized or not—should be based on the results of the primary outcome measure as analyzed in a prespecified statistical analysis plan.

Conclusions from observational studies should be based on the results of multivariable models or other methods that control for correlated data, potential confounding, and effect modification. Conclusions should not be based on

unadjusted analyses with a single predictor (independent or explanatory variable) unless confounding can be excluded.

Closing

Clinical research is difficult, and truth is elusive. The best that can be done is to conduct a well-designed study as rigorously as possible, to acknowledge its shortcomings, to present the results fairly, and to interpret treatment effects carefully, neither overstating their importance nor understating uncertainty (Pollock 2020).

Evidence-based dentistry is literature-based dentistry (Lang 2010). Clinicians, authors, reviewers, and editors should take the time to learn how to accurately report and assess the validity, relevance, and implications of the published literature. The Cochrane Center is the premier site for systematic reviews (The Cochrane Collaboration n.d.). Sites such as the ADA Center for Evidence-Based Dentistry (Center for Evidence-Based Medicine n.d.) and the University of Dundee Centre for Evidence-Based Dentistry (University of Dundee, School of Dentistry n.d.) make it easy to find clinical guidelines. Such clinical guidelines depend directly on the existing evidence and on the ability to appraise that evidence.

Ultimately, patient care is improved when valid and useful research is planned, executed, and communicated to practitioners. The guidelines presented here should assist authors in preparing research reports, journal editors in reviewing those reports, and clinicians in understanding those reports. Journal editors can also disseminate these guidelines by including them in their instructions to authors and insisting that authors follow them as a condition of publication. We also hope the OHStat Guidelines will serve as a template for updating and informing increasingly useful oral health research reports.

Author Contributions

A.M. Best and T.A. Lang contributed to the conception and design of the

guidelines, took the lead in organizing, drafting, and documenting the original manuscript, and incorporated comments and insights from the other authors. J.C. Gunsolley, E. Ioannidou, and B.L. Greenberg contributed to the conception of the guidelines, critically appraised each revision, and provided substantive comments and insights throughout the development process. All authors agree to be accountable for all aspects of the work and approved the final draft for publication.

Disclaimer

This article was written by the Task Force writing group authors, who take sole responsibility for the final content. The views expressed do not represent the policies, views, or opinions of the authors' institutions.

Task Force Writing Group

A.M. Best, Virginia Commonwealth University; B.L. Pihlstrom, University of Minnesota; D.V. Dawson, University of Iowa; B.L. Greenberg, New York Medical College; E. Ioannidou, University of Connecticut; J.C. Gunsolley, Virginia Commonwealth University; J.S. Hodges, University of Minnesota; and T.A. Lang, Principal, Tom Lang Communications and Training International. P.B. Imrey, Cleveland Clinic and Case Western Reserve University, also provided suggested revisions and performed comprehensive reviews.

Reviews

Substantial written critiques were also provided by M. Glick, *JADA* Editor; N.S. Jakobovics, *JDR* Editor; W.M. Thomson, *Community Dentistry and Oral Epidemiology* Editor; B.S. Michalowicz, University of Minnesota; L.N. Borrell, City University New York; J.M. Grender and J. DiGennaro, Procter and Gamble; M. Hayat, Georgia State University; R. Brignardello-Petersen, McMaster University; P. Imrey, Cleveland Clinic and Case Western; D. Wu, University of North Carolina; P.G. Robinson and T. Walsh, *Community Dental Health Journal*. We gratefully acknowledge their contributions; the final manuscript does

not necessarily reflect either their views or the official policy of the organizations they represent.

Declaration of Conflicting Interests

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: All authors have completed the ICMJE unified competing interest form, a copy of which is available from the corresponding author.

To encourage dissemination of the OHStat Statement, this article and the checklist is freely available on <http://taskforceondesign.org/>. This article has been simultaneously copublished in the *JDR Clinical and Translational Research*, *The Journal of the American Dental Association*, and the *Journal of Oral and Maxillofacial Surgery*. The articles are identical except for minor stylistic and spelling differences in keeping with each journal's style. Any of those journal's citations can be used when citing this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The Task Force on Design and Analysis in Oral Health Research provided funding for the December 2020 meeting and provided support for the consultant (T.L.). A.B. received funding for travel related to the December 2020 meeting. None of the Task Force sponsors was involved in the planning, execution, or writing of the OHStat documents. Additionally, no funder played a role in the drafting of the manuscript.

ORCID iD

A.M. Best  <https://orcid.org/0000-0002-6300-3925>

References

- Abraham I, Montedori A. 2010. Modified intention to treat reporting in randomised controlled trials: systematic review. *BMJ*. 340:c2697.
- Albandar JM, Goldstein H. 1992. Multi-level statistical models in studies of periodontal diseases. *J Periodontol*. 63(8):690–695.

- Ali Z, Baker S, Sereno N, Martin N. 2020. A pilot randomized controlled crossover trial comparing early OHRQoL outcomes of cobalt-chromium versus PEEK removable partial denture frameworks. *Int J Prosthodont.* 33(4):386–392.
- Altman DG, Bland JM. 1995. Statistics notes: absence of evidence is not evidence of absence. *BMJ.* 311(7003):485.
- Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, Gotzsche PC, Lang T. 2001. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med.* 134(8):663–694.
- American Dental Association (ADA). 2013. ADA Clinical Practice Guidelines Handbook. (November):58. [accessed 2021 Jun 30]. http://ebd.ada.org/~media/EBD/Files/ADA_Clinical_Practice_Guidelines_Handbook-2013.pdf.
- Bastian H. 2014. Science in the abstract: don't judge a study by its cover. *Scientific American.* 12 May.
- Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, et al. 1996. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA.* 276(8):637–639.
- Bellamy SL, Lin JY, Ten Have TR. 2007. An introduction to causal modeling in clinical trials. *Clin Trials.* 4(1):58–73.
- Best AM, Greenberg B, Glick M. 2016. From tea tasting to t test: a P value ain't what you think it is. *J Am Dent Assoc.* 147(7):527–529.
- Best AM, Lang TA, Greenberg BL, Gunsolley JC, Ioannidou E, Task Force on Design and Analysis in Oral Health Research, 2024. The OHSat Guidelines for reporting observational studies and clinical trials in oral health research: manuscript checklist. *J Dent Res.* In press.
- Biomarkers Definitions Working Group. 2001. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther.* 69(3):89–95.
- Blance A, Tu YK, Baelum V, Gilthorpe MS. 2007. Statistical issues on the analysis of change in follow-up studies in dental research. *Commun Dent Oral Epidemiol.* 35(6):412–420.
- Brignardello-Petersen R, Carrasco-Labra A, Shah P, Azarpazhooh A. 2013. A practitioner's guide to developing critical appraisal skills: what is the difference between clinical and statistical significance? *J Am Dent Assoc.* 144(7):780–786.
- Brody T. 2016. Intent-to-treat analysis versus per protocol analysis. In: *Clinical Trials.* 2nd ed. Elsevier. p. 173–201.
- Casamassimo PS, Thikkurissy S, Edelstein BL, Maiorini E. 2009. Beyond the dmft: the human and economic cost of early childhood caries. *J Am Dent Assoc.* 140(6):650–657.
- Center for Evidence-Based Medicine. n.d. Teaching materials. [accessed 2021 Jan 2]. <https://www.cebm.org/teaching-materials/>.
- Chalkidou K, Tunis S, Whicher D, Fowler R, Zwarenstein M. 2012. The role for pragmatic randomized controlled trials (pRCTs) in comparative effectiveness research. *Clin Trials.* 9(4):436–446.
- Chilton NW, Fleiss JL. 1986. Design and analysis of plaque and gingivitis clinical trials. *J Clin Periodontol.* 13(5):400–406.
- Christiansen S, Iverson C, Flanagan A, Fontanarosa P, Glass R, Glitman P, Lantz J, Meyer H, Smith J, Winker M, et al. 2020. *AMA Manual of Style: A Guide for Authors and Editors.* 11th ed. New York (NY): Oxford University Press.
- Christogiannis C, Nikolakopoulos S, Pandis N, Mavridis D. 2022. The self-fulfilling prophecy of post-hoc power calculations. *Am J Orthodont Dentofacial Orthop.* 161(2):315–317.
- Council of Science Editors. 2015. *Scientific Style and Format: The CSE Manual for Authors, Editors, and Publishers.* 8th ed. Chicago (IL): University of Chicago Press.
- De Castro P, Heidari S, Babor TF. 2016. Sex And Gender Equity in Research (SAGER): reporting guidelines as a framework of innovation for an equitable approach to gender medicine. *Commentary. Ann Ist Super Sanita.* 52(2):154–157.
- DeRouen TA. 1990. Statistical models for assessing risk of periodontal disease. In: Bader JD, editor. *Risk Assessment in Dentistry.* Chapel Hill: University of North Carolina. p. 239–244.
- DeRouen TA, Mancl L, Hujuel P. 1991. Measurement of associations in periodontal diseases using statistical methods for dependent data. *J Periodont Res.* 26(3):218–229.
- Dorsey R, Graham G. 2011. New HHS data standards for race, ethnicity, sex, primary language, and disability status. *JAMA.* 306(21):2378–2379.
- Ebbutt AF, Frith L. 1998. Practical issues in equivalence trials. *Stat Med.* 17(15–16):1691–1701.
- Eke PI, Thornton-Evans GO, Wei L, Borgnakke WS, Dye BA. 2010. Accuracy of NHANES periodontal examination protocols. *J Dent Res.* 89(11):1208–1213.
- Erasmus A, Holman B, Ioannidis JPA. 2022. Data-dredging bias. *BMJ Evid Based Med.* 27(4):209–211.
- Etminan M, Brophy JM, Collins G, Nazemipour M, Mansournia MA. 2021. To adjust or not to adjust: the role of different covariates in cardiovascular observational studies. *Am Heart J.* 237:62–67.
- Fleiss JL, Wallenstein S, Chillton NW, Goodson JM. 1988. A re-examination of within-mouth correlations of attachment level and of change in attachment level. *J Clin Periodontol.* 15(7):411–414.
- Fleming TR, Powers JH. 2012. Biomarkers and surrogate endpoints in clinical trials. *Stat Med.* 31(25):2973–2984.
- Flight L, Julious SA. 2016. Practical guide to sample size calculations: non-inferiority and equivalence trials. *Pharm Stat.* 15(1):80–89.
- Friedman LM, Furberg CD, DeMets DL, Reboussin DM, Granger CB, Friedman LM, Furberg CD, DeMets DL, Reboussin DM, Granger CB. 2015. The randomization process. In: *Fundamentals of Clinical Trials.* p. 123–145.
- Gamble C, Krishan A, Stocken D, Lewis S, Juszcak E, Doré C, Williamson PR, Altman DG, Montgomery A, Lim P, et al. 2017. Guidelines for the content of statistical analysis plans in clinical trials. *JAMA.* 318(23):2337–2343.
- Gelman A, Loken E. 2014. The statistical crisis in science. *Am Scientist.* 102(6):460–465.
- Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, Debeer H, et al. 2011. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol.* 64(4):383–394.
- Hartman JM, Forsen JW, Wallace MS, Neely JG. 2002. *Tutorials in clinical research: Part IV: recognizing and controlling bias.* *Laryngoscope.* 112(1):23–31.
- Heidari S, Babor TF, De Castro P, Tort S, Curno M. 2016. Sex and Gender Equity in Research: rationale for the SAGER guidelines and recommended use. *Res Integrity Peer Review.* 1(1):2.
- Hill B. 1965. The reasons for writing. *BMJ.* 9 October:870.
- Holtfreter B, Alte D, Schwahn C, Desvarieux M, Kocher T. 2012. Effects of different manual periodontal probes on periodontal

- measurements. *J Clin Periodontol.* 39(11):1032–1041.
- Hornberger J, Wrona E. 1997. When to base clinical policies on observational versus randomized trial data. *Ann Intern Med.* 127(8, Suppl. II):697–703.
- Hsieh FY. 1989. Sample size tables for logistic regression. *Stat Med.* 8(7):795–802.
- Hujoel PP. 1998. Design and analysis issues in split mouth clinical trials. *Commun Dent Oral Epidemiol.* 26(2):85–86.
- Hujoel PP, Loesche WJ. 1990. Efficiency of split-mouth designs. *J Clin Periodontol.* 17(10):722–728.
- Hujoel PP, Moulton LH. 1988. Evaluation of test statistics in split-mouth clinical trials. *J Periodont Res.* 23(6):378–380.
- Hyman J. 2006. The importance of assessing confounding and effect modification in research involving periodontal disease and systemic diseases. *J Clin Periodontol.* 33(2):102–103.
- Imrey PB. 1986. Considerations in the statistical analysis of clinical trials in periodontitis. *J Clin Periodontol.* 13(5):517–532.
- Imrey PB, Chilton NW, Pihlstrom BL, Proskin HM, Kingman A, Listgarten MA, Zimmerman SO, Ciancio SG, Cohen ME, D'Agostino RB. 1994. Proposed guidelines for American Dental Association acceptance of products for professional, non-surgical treatment of adult periodontitis. Task Force on Design and Analysis in Dental and Oral Research. *J Periodont Res.* 29(5):348–360.
- International Committee of Medical Journal Editors. 2018. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. Vancouver: International Committee of Medical Journal Editors.
- Ioannidis JPA, Evans SJW, Gøtzsche PC, O'Neill RT, Altman DG, Schulz K, Moher D. 2004. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med.* 141(10):781–788.
- Katz BP. 1997. Measuring quality, outcomes, and cost of care of using large databases: preface. *Ann Intern Med.* 127(8, Suppl. II):665.
- Lachin JM. 2000. Statistical considerations in the intent-to-treat principle. *Control Clin Trials.* 21(3):167–189.
- Lang T. 2018. Up and down or side by side: structuring comparisons in data tables. *AMWA J.* 33(3):104–110.
- Lang TA. 2010. How to Write, Publish, and Present in the Health Sciences: A Guide for Clinicians and Laboratory Researchers. ACP Press.
- Lang TA. 2017. Writing a better research article. *J Public Health Emerg.* 1:88–88.
- Lang TA. 2020. An author's editor reads the "instructions for authors." *Eur Sci Editing.*
- Lang TA. 2022. Scientific abstracts: texts, contexts, and subtexts. *Eur Sci Editing.* 48:e85616.
- Lang T, Altman DG. 2013. Basic statistical reporting for articles published in clinical medical journals: the SAMPL Guidelines. In: Smart P, Maisonneuve H, Polderman A, eds. *Science Editors' Handbook.* European Association of Science Editors.
- Lang TA, Secic M. 2006. How to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors, and Reviewers. 2nd ed. American College of Physicians.
- Lang TA, Stroup DF. 2020. Who knew? The misleading specificity of "double-blind" and what to do about it. *Trials.* 21(1):697.
- Lang TA, Talerico C, Siontis GCM. 2012. Documenting clinical and laboratory images in publications: the CLIP principles. *Chest.* 141(6):1626–1632.
- Lesaffre E, Philstrom B, Needleman I, Worthington H. 2009. The design and analysis of split-mouth studies: what statisticians and clinicians should know. *Stat Med.* 28:3470–3482.
- Masood Y, Bower P, Waheed MW, Brown G, Waheed W. 2019. Synthesis of researcher reported strategies to recruit adults of ethnic minorities to clinical trials in the United Kingdom: a systematic review. *Contemp Clin Trials.* 78:1–10.
- McDonald BW, Pack ARC. 1990. Concepts determining statistical analysis of dental data. *J Clin Periodontol.* 17(3):153–158.
- Mills JL. 1993. Data torturing. *N Engl J Med.* 329:1196–1199.
- Moher D, Jones A, Lepage L. 2001. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA.* 285(15):1992–1995.
- Nguyen TV, Rivadeneira F, Civitelli R. 2019. New Guidelines for data reporting and statistical analysis: helping authors with transparency and rigor in research. *J Bone Miner Res.* 34(11):1981–1984.
- Nørskov AK, Lange T, Nielsen EE, Gluud C, Winkel P, Beyersmann J, De Uná-Álvarez J, Torri V, Billot L, Putter H, et al. 2021. Assessment of assumptions of statistical analysis methods in randomised clinical trials: the what and how. *BMJ Evid Based Med.* 26(3):121–126.
- Office of the Secretary. 2017. ORI Activity Summaries. *Office Res Integrity Newsl.* 24(1):7–8.
- Oxford for Evidence-Based Medicine Group. 2013. The Oxford Levels of Evidence 2. The Oxford Centre for Evidence-Based Medicine. <http://www.cebm.net/index.aspx?o=5653>.
- Pandis N. 2012. Sample calculation for split-mouth designs. *Am J Orthodont Dentofacial Orthop.* 141(6):818–819.
- Pandis N, Chung B, Scherer RW, Elbourne D, Altman DG. 2019. CONSORT 2010 statement: extension checklist for reporting within person randomised trials. *Br J Dermatol.* 180(3):534–552.
- Pandis N, Polychronopoulou A, Eliades T. 2010. An assessment of quality characteristics of randomised control trials published in dental journals. *J Dent.* 38(9):713–721.
- Pandis N, Polychronopoulou A, Madianos P, Makou M, Eliades T. 2011. Reporting of research quality characteristics of studies published in 6 major clinical dental specialty journals. *J Evid Based Dent Pract.* 11(2):75–83.
- Piaggio G, Elbourne DR, Pocock SJ, Evans SJW, Altman DG. 2012. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA.* 308(24):2594–2604.
- Pihlstrom BL, Barnett ML. 2010. Design, operation, and interpretation of clinical trials. *J Dent Res.* 89(8):759–772.
- Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, Gaboury I. 2006. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust.* 185(5):263–267.
- Pocock SJ, Hughes MD, Lee RJ. 1987. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med.* 317(7):426–432.
- Pollock NW. 2020. Managing bias in research. *Wilderness Environ Med.* 31(1):1–2.
- Pulit SL, de With SAJ, de Bakker PIW. 2017. Resetting the bar: statistical significance in whole-genome sequencing-based association studies of global populations. *Genet Epidemiol.* 41(2):145–151.
- Qin D, Hua F, He H, Liang S, Worthington H, Walsh T. 2020. Quality of split-mouth trials in dentistry: 1998, 2008, and 2018. *J Dent Res.* 99(13):1453–1460.

- Qu H-Q, Tien M, Polychronakos C. 2010. Statistical significance in genetic association studies. *Clin Invest Med*. 33(5):E266–E270.
- Rossner M, Yamada KM. 2004. What's in a picture? The temptation of image manipulation. *J Cell Biol*. 166(1):11–15.
- Rozshart JJ, Kumar SS, Allareddy V, Childs CA, Elangovan S. 2020. Spin in abstracts of randomized controlled trials in dentistry: a cross-sectional analysis. *J Am Dent Assoc*. 151(1):26–32.e3.
- Sackett DL. 2004. Turning a blind eye: why we don't test for blindness at the end of our trials. *BMJ*. 328(7448):1136.
- Schulz KF, Altman DG, Moher D. 2010. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *Ann Intern Med*. 152(11):736–732.
- Schulz KF, Chalmers I, Hayes RJ, Altman DG. 1995. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*. 273(5):408–412.
- Shadish WR, Cook TD, Campbell DT. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. New York (NY): Houghton, Mifflin and Company.
- Shaqman M, Al-Abedalla K, Wagner J, Swede H, Gunsolley JC, Ioannidou E. 2020. Reporting quality and spin in abstracts of randomized clinical trials of periodontal therapy and cardiovascular disease outcomes. *PLoS ONE*. 15(4):e0230843.
- Shell ER. 2016. Hurdling obstacles. *Science*. 353(6295):116–119.
- Shih W. 2002. Problems in dealing with missing data and informative censoring in clinical trials. *Curr Control Trials Cardiovasc Med*. 3(1):4.
- Shrier I, Steele RJ, Verhagen E, Herbert R, Riddell CA, Kaufman JS. 2014. Beyond intention to treat: what is the right question? *Clin Trials*. 11(1):28–37.
- Smiley CJ, Tracy SL, Abt E, Michalowicz BS, John MT, Gunsolley J, Cobb CM, Rossmann J, Harrel SK, Forrest JL, et al. 2015. Evidence-based clinical practice guideline on the nonsurgical treatment of chronic periodontitis by means of scaling and root planing with or without adjuncts. *J Am Dent Assoc*. 146(7):525–535.
- Smith PJ, Hadgu A. 1992. Sensitivity and specificity for correlated observations. *Stat Med*. 11(11):1503–1509.
- Sterne JAC, Johnson NW, Wilton JMA, Joyston-Bechal S, Smales FC. 1988. Variance components analysis of data from periodontal research. *J Periodont Res*. 23(2):148–153.
- Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, Henry D, Altman DG, Ansari MT, Boutron I, et al. 2016. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 355:i4919.
- Steyerberg EW, Harrell FEJ. 2016. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*. 69:245–247.
- Talbot D, Massamba VK. 2019. A descriptive review of variable selection methods in four epidemiologic journals: there is still room for improvement. *Eur J Epidemiol*. 34(8):725–730.
- Task Force on Design. n.d. Task Force on Design and Analysis of Oral Health Research. [accessed 2021 May 7]. taskforceondesign.org.
- The Cochrane Collaboration. n.d. About us. [accessed 2021 Jan 7]. <https://www.cochrane.org/>
- Torabinejad M, Bahjri K. 2005. Essential elements of evidenced-based endodontics: steps involved in conducting clinical research. *J Endod*. 31(8):563–569.
- University of Dundee, School of Dentistry. n.d. Dundee S. Centre for Evidence Based Dentistry. [accessed 2021 Oct 2]. <https://www.cebd.org/>
- University of Oxford Center for Statistics in Medicine. n.d. EQUATOR (Enhancing the QUALity and Transparency Of health research) network. [accessed 2021 Jan 7]. <https://www.equator-network.org/>
- U.S. National Institutes of Health. 2014. Frequently asked questions: NIH clinical trial definition. [accessed 2021 Oct 2]. <https://grants.nih.gov/faqs#/clinical-trial-definition.htm>
- Vickers AJ. 2016. Change/percent change from baseline. In: *Wiley StatsRef: Statistics Reference Online*. Chichester (UK): John Wiley. p. 1–7.
- von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. 2014. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Int J Surg*. 12(12):1495–1499.
- Wainer Z, Carcel C, Hickey M, Schiebinger L, Schmiede A, McKenzie B, Jenkins C, Webster J, Woodward M, Hehir A, et al. 2020. Sex and gender in health research: updating policy to reflect evidence. *Med J Aust*. 212(2):57–62.e1.
- Wasserstein RL, Lazar NA. 2016. The ASA statement on p-values: context, process, and purpose. *Am Stat*. 70(2):129–133.
- Wasserstein RL, Schirm AL, Lazar NA. 2019. Moving to a world beyond “ $p < 0.05$.” *Am Stat*. 73(Suppl 1):1–19.
- Wood AM, White IR, Thompson SG. 2004. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clin Trials*. 1(4):368–376.
- World Association of Medical Editors. n.d. WAME statement on conflict of interest in peer-reviewed medical journals. 2009. [accessed 2023 Jan 9]. <https://www.wame.org/conflict-of-interest-in-peer-reviewed-medical-journals>.
- World Health Organization. 2014. *International Clinical Trials Registry Platform*. Search Portal. Geneva (Switzerland): WHO.
- Zhang C, Liu X. 2011. Review of James Hartley's research on structured abstracts. *J Inform Sci*. 37(6):570–576.