Editorial

Statistical Reviewing for *Headache*

Timothy T. Houle, PhD; Donald B. Penzien, PhD

The process of reviewing the statistical analyses in a manuscript is a daunting one. There is an ever-changing list of statistical tests, procedures, and best practices. The available statistical software to conduct analyses is similarly changing and many reviewers are probably finding it more and more difficult to interpret the presented data in a manuscript. Perhaps it is no wonder, then, that reviewers often neglect to comment on the statistical analyses of a manuscript, or provide comments that do not serve to improve the scientific product. The difficulties in statistical reviewing have been well documented,[1] but a looming question remains: What can our research community do about improving the statistical aspects of peer review?

Examination of publication practices reveal that changes in the behavior of reviewers can be difficult to initiate and maintain. Nevertheless, efforts at enhancing statistical peer review must be made if we are to see improvements in this important reviewing issue. The editorial staff of *Headache* has developed a new set of reviewer guidelines to assist reviewers in formulating constructive criticisms of submitted manuscripts. These guidelines also can provide a valuable resource for authors as they work to prepare manuscripts for submission. The purpose of this editorial is to introduce a statistical reviewing checklist that is embedded within these guidelines. The statistical checklist is intended to serve as an initial step in assisting reviewers for *Headache* to formulate basic criticisms of the statistical reporting and design of submitted manuscripts.

REVIEWING STATISTICAL REPORTING

In assessing the statistical quality of a manuscript, a reviewer can prudently focus on one overarching review question: Have the statistical methods been presented in sufficient detail such that they could be replicated? Too often, statistical methods are presented in insufficient detail. This leads to a scenario wherein a reader cannot focus on the presented data because of uncertainty that the methods were properly conducted. Good statistical reporting allows the actual data to be the focus of the manuscript. Table 1 presents basic guidelines that, if satisfied, will lead to improved statistical reporting. Reviewers are encouraged to submit inquiries to authors if any of these reporting issues have not been adequately addressed.

Reporting the results of statistical tests is a crucial element of statistical reporting. Too often, authors rely on reporting *P* values outside the context of information concerning the effect sizes of the observed differences.[2,3] Relying too heavily on significance testing may not allow a manuscript to reach its full potential, for to properly interpret a *P* value many other statistical reporting elements must also be presented. Figure 1 provides a simple guiding principle regarding the information value of statistical results based on the question: What do these differences (effects) actually convey? Addressing this question through good reporting is crucial as a statistically significant effect can accrue in the absence of a large

From the Department of Anesthesiology, Wake Forest University Medical School, Winston-Salem, NC, USA (T.T. Houle); Head Pain Center, University of Mississippi Medical Center, Jackson, MS, USA (D.B. Penzien)

Table 1.—Reviewing (Preparing) Statistical Reporting in *Headache*

Category	Review question	Notes
Statistical software used to conduct the analysis	What software is used?	EXCEL, SAS, SPSS, SYSTAT, R, etc.
Significance level and nature of the hypothesis	How is statistical significance determined?	Type I error level at $\alpha = 0.05$ by convention. One or two-tailed testing, equivalence, inferiority, etc.
Exact *P* values	What is the exact *P* value?	Preferable to "$P < .05$"
Descriptive statistics	Are descriptive statistics optimally reported?	Appropriate to quality/level of data. Mean and standard deviation are recommended for distributions that satisfy parametric assumptions, while median and interquartile range (IQR) are recommended for skewed data or data with outliers
Analytical plan	How are the data analyzed?	Not just a list of tests. Indicates which variables were analyzed by which tests, error control strategy, rationale for choice of tests, missing data strategy (eg, intent-to-treat), etc.
Subject/data disposition	How many subjects are in the analysis and is there a potential for bias based on missing data?	Extent of missing data, available sample size, outlier (artifact) characterization, alternative samples, etc.
Statistical results	What does the magnitude of the observed differences indicate?	Not just *P* values reported; see Figure 1.

effect size (eg, treatment effect), and even large treatment effects may not have important clinical significance.

REVIEWING STATISTICAL DESIGNS

A more difficult issue to review is the design of statistical analyses or the analytical plan. In this regard, peer review does not necessarily result in sound scientific research,[4] and even well-intentioned reviewers may submit mistaken criticism.[5] Reviewers come to the peer review process with vast differences in knowledge of research design and statistical practices. Because of this, *Headache* employs the services of a statistical consult to assist in the review process for articles that have been identified as needing additional statistical review. *Headache* is not alone in this policy, but like many other medical journals with similar levels of circulation, formal statistical review
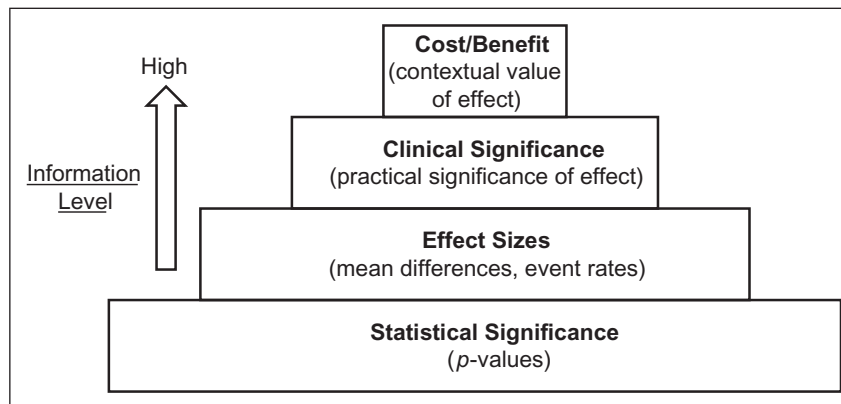

Fig 1.—An information hierarchy for statistical reporting. Authors should be encouraged to report the highest level of information available. Statistical significance (*P*-values) judgments are not nearly as informative as when also presented with a measure of effect size. When available, the clinical (practical) significance of findings and even cost/benefit ratios of observed effects provide a wealth of information to readers.

Table 2.—Reviewing (Preparing) Statistical Designs in *Headache*

Category	Review question	Notes
Quality/level of data matches applied statistical test	Does the level/quality of data allow for the utilized test(s)?	Nominal or categorical data analyzed using appropriate tests; ordinal (ie, rank-order) data analyzed using nonparametric tests
Parametric assumptions	Have parametric assumptions been evaluated for parametric tests?	Although many tests are robust to violations in their underlying assumptions, violations of parametric assumptions (ie, normality, homogeneity of variances) should be considered
Specific test assumptions	Are these tests appropriate for these data?	Every test assumes something specific about the data being analyzed. For example, an independent *t*-test assumes that each subject contributes only one score to the analysis
Rejecting the null hypothesis	Does the choice of statistical test(s) allow the authors to properly evaluate the hypotheses?	Every test examines a specific null-hypothesis that may or may not support the claims made by the authors. For example, comparing the *P* values of two *t*-tests is not equivalent to an interaction effect in ANOVA

presently does not occur on every submitted manuscript.[6] It is for this reason that guidelines for the review of statistical designs are proposed to assist our reviewers in formulating their criticism.

Table 2 presents several review questions recommended to guide the review of a submitted analytical plan. Reviewing the analytical plan doubtless requires substantial knowledge about the statistical tests under consideration, and in particular, the assumptions underlying the tests. Many reviewers may opt to defer review on the statistical design of a study. However, the headache research community consists of a wealth of talented methodologists/ clinicians and researchers who routinely conduct and evaluate research, and it is for these reviewers that the checklist is designed.

GUIDELINES FOR REVIEWERS

The newly developed guidelines for reviewers can be found on the Journal web site (http:// www.headachejournal.org), at the login page for the manuscript submission and review (http:// mc.manuscriptcentral.com/headache) and as an attachment to the template e-mail dispatched when a

reviewer agrees to provide a review. This editorial is meant to serve as an introduction to the statistical checklists therein.

Editor's Note: Dr. Houle serves as the statistical consultant for Headache: The Journal of Head and Face Pain.

REFERENCES

1. Altman DG. Statistical reviewing for medical journals. *Stat Med*. 1998;17:2661-2674.
2. Houle TT. Statistical reporting for current and future readers. *Anesthesiology*. 2007;107:193-194.
3. Houle TT. Importance of effect sizes for the accumulation of knowledge. *Anesthesiology*. 2007;106:415-417.
4. Altman DG. Poor-quality medical research: What can journals do? *JAMA*. 2002;287:2765-2767.
5. Bacchetti P. Peer review of statistics in medical research: The other problem. *BMJ*. 2002;324:1271-1273.
6. Goodman SN, Altman DG, George SL. Statistical reviewing policies of medical journals: Caveat lector? *J Gen Intern Med*. 1998;13:753-756.