Journal of Traumatic Stress, Vol. 11, No. 3, 1998

Statistical Review: An Approach to Common Methodological and Statistical Problems

Paula P. Schnurr[1,2]

The field of traumatic stress studies has grown significantly since the *Journal of Traumatic Stress* first began publication in 1988. Accompanying this growth has been an increase in the complexity of traumatic stress research, especially with respect to the widespread availability of advanced methods for addressing complicated statistical problems. To keep pace with these changes, the *Journal* has recently instituted statistical review procedures. This article describes the review process and provides some guidelines to assist authors in attending to the methodological and statistical aspects of their manuscripts.

Most manuscripts submitted to the *Journal of Traumatic Stress* will be statistically reviewed at the point in the content (regular) review process when the Editor feels the manuscript is likely to be accepted, but the review may occur sooner if the Editor feels that statistical review at an earlier stage is necessary in order to make a decision. Selected manuscripts will be statistically reviewed as part of the initial content review process. Generally, these selected manuscripts will be those that report on advanced methods such as structural equation modeling, meta-analysis, hierarchical linear models, and signal detection analysis.

Statistical reviewers are individuals from the fields of psychiatry, psychology, epidemiology, and biostatistics who have demonstrated quantitative expertise. Many are trauma researchers, although some are not. A number of the present statistical reviewers have been content reviewers at the *Journal of Traumatic Stress* and continue to serve in that capacity. Readers interested in serving as a statistical reviewer or in recommending a colleague as a reviewer should contact the author.

[1]Department of Veterans Affairs and Department of Psychiatry, Dartmouth Medical School.
[2]To whom correspondence should be addressed at the VA National Center for PTSD (116D), VA Medical and Regional Office Center, White River Junction, Vermont 05009.

0894-9867/98/0700-0405$15.00/1 © 1998 International Society for Traumatic Stress Studies

The Content of Statistical Review

The questions addressed by the statistical reviewers are modeled on the set of statistical review questions developed by Drs. John Bartko and Lee Gurel at the *American Journal of Psychiatry*. Questions used at the *Journal of Traumatic Stress* are listed below, along with information about common problems in each area. It is strongly recommended that authors review these questions as a guide for preparing their manuscripts and as a checklist immediately prior to submission.

PURPOSE: Are the Hypotheses or Objectives Clearly Stated?

All too often, authors fail to clarify what question or questions a study is designed to ask. Hypotheses, too, often are not stated. This information is essential for evaluating the design, method, and results and should be included in the introduction to every manuscript.

DESIGN: Is the Design Appropriate for Testing the Hypotheses or Meeting the Objectives, e.g., Are the Control Groups Appropriate?

There are few hard and fast rules about what is a "good" versus a "bad" design. The real issue is whether a given design supports the testing of a given hypothesis, or permits the inferences that authors wish to draw. For example, in a case-control design comparing combat veterans with PTSD to nonpsychiatric veteran controls who have not been in combat, little can be said about the specificity of findings to PTSD; any difference between PTSD and no PTSD groups may be due to the fact that the men with PTSD have a psychiatric disorder and not to the PTSD itself. Another design issue is the nature of the comparison treatment in a psychotherapy treatment study. A randomly assigned wait-list control group rules out many threats to internal validity and is appropriate for research that is at an early stage of investigation. A more active comparison group is needed to establish construct validity—why a particular treatment works—and should be used when evaluating an established treatment. For guidance, a chapter by Borkovec (1993) offers detailed discussion of methodological issues in psychotherapy research. Another useful volume is the book on quasi-experimental design by Cook and Campbell (1979), relevant because of the non-experimental nature of most studies in the field of traumatic stress. This book can be valuable for determining what factors might threaten the validity of a study and what inferences might be drawn in light of these threats.

SAMPLE: Is the Sample Adequately Described?

It is important to describe the demographic (and, when relevant, psychiatric) status of participants so that readers can evaluate the generalizability of findings and decide whether the results are relevant to their interests. Besides incomplete descriptive information, another common problem is the information about sample size, which is admittedly a difficult area when one is trying to honestly present information about attrition or is describing subgroups for secondary analyses. Nevertheless, most readers have experienced the frustration of reading a method section several times trying to figure out exactly how many participants were in a particular study.

RECRUITMENT/PARTICIPATION: Is There Adequate Information About How Participants Were Recruited? Is There Adequate Information About Nonparticipants? Are Recruitment Procedures Adequate? Are These Issues Discussed and Handled Appropriately?

There are two general issues being considered in this set of questions: (1) whether the recruitment procedures resulted in a representative sample and minimized threats to internal validity and (2) whether the processes of recruitment and selection are fully described. Almost all studies reflect some degree of nonparticipation or attrition, so it is virtually always necessary to provide readers with sufficient information for them to evaluate threats to validity. Furthermore, it is sometimes necessary to address nonparticipation/attrition in statistical analysis (e.g., by using ANCOVA because differential dropout caused pretreatment scores for treatment and placebo groups to differ), as well as in discussion, by noting how the conclusions may be limited.

POWER: Is the Sample Size Adequate for Testing the Proposed Hypotheses or Meeting the Proposed Objectives? Is the Discussion of Power Issues Adequate?

The answer to these questions is "no" far too often. The problem is not specific to the field of traumatic stress studies by any means, but to much research in general. We ask that authors address issues of power and effect size to facilitate the interpretation of null findings. For example, you may wish to conclude that two treatments do not differ because there was no statistically significant ($p < .05$, two-tailed) difference in outcome between them using samples of 18 people each treatment group. Yet with a

sample of this size, power is only sufficient (>80%, a conventional minimum) to find a very large treatment effect—a difference (d) of 1.0 or more standard deviations! If the true effect were $d = .5$, which is considered to be medium (Cohen, 1988), the probability of Type II error would be 69%. In such a case, the appropriate inference is that the treatments are unlikely to differ *by a large amount*. Also, the possibility of Type II error needs to be discussed and ruled out as a threat to validity if smaller effects would be expected. Null results in a study that did not have adequate power to find meaningful results may compromise the likelihood of publication. The volume on power analysis by Cohen (1988) is recommended to facilitate both the interpretation and the planning of studies.

MEASUREMENT: Is There Adequate Information About Measurement? Are Reliability and Validity Issues Adequately Handled?

The amount of detail to report about the measures used in a study varies depending on the availability of relevant information about the psychometric properties of the measures. In the extreme case, where authors have developed a measure for a particular study, information about reliability and validity needs to be provided. If this information is unavailable, the effect on the interpretation of findings needs to be addressed in the discussion. Information about interrater reliability is useful for interview-based measures and generally should be reported as kappa or an intraclass correlation coefficient, depending on the scale of measurement.

DATA ANALYSIS: Are the Statistical Tests Appropriate for Testing the Proposed Hypotheses or Meeting the Proposed Objectives? Are the Data Analyses Well Done? Is All Conventionally Reported Information Included?

Frequently, the data analyses of a potentially important study are poorly done, thereby limiting the inferences that can be drawn from the findings. A related problem is that analyses reported do not test the proposed hypotheses. There are far too many specific problems, and solutions, to mention them here. An incomplete list of some common problems includes: too few cases relative to the number of variables in a multivariate analysis, inadequate control for Type I error, failure to meet assumptions of a given test, and incorrect parametrization for tests of interactions or contrasts in a linear model. Instead, readers are referred to the book by Tabachnick and Fidell (1996), a clear and comprehensive text with excellent

examples of how to format tables and describe the results of various types of statistical procedures. A more focused book by Aiken and West (1991) also is recommended because it addresses a difficult and problematic topic—testing interactions in multiple regression. Readers are asked to attend to details as well, by reporting complete statistical test information for all statistically significant results and for important nonsignificant results as well. "Complete" means p-values *and* degrees of freedom for each test value, as well as test-specific conventions such as the overall r^2 for multiple least squares regression and the prior probabilities assumed for classification in logistic regression and discriminant analysis.

MISSING DATA: Are Missing Data Handled Appropriately? Is Variation in Sample Size From Analysis to Analysis Understandable and Appropriate?

This item refers to the quantitative procedures used to handle missing data. There are almost always missing data in a study, especially if the study is longitudinal. In general, the sample size for a study should be defined as the number of participants who have complete data on the measures of primary interest. It is confusing for readers to find sample size varying from analysis to analysis. Furthermore, comparisons across outcomes may be invalid if there are large variations in sample size across outcomes, especially if the likelihood of missing data is related to participant characteristics or outcomes. If analyses of subsets of participants are conducted, any systematic variation in the characteristics of the subsets compared with the remaining participants needs to be addressed by statistical analysis (e.g., showing that the subset does not differ from the remaining sample on important outcome or predictor variables) or by logical argument. For longitudinal analysis, the practice known as "last endpoint carried forward," in which a participant's score at the point of dropout is used as the value for all subsequently missing data, is inappropriate. A better approach is to design the primary analyses to retain the maximum number of participants and then to treat analyses with smaller sample sizes as secondary. Sometimes it may be possible to use a longitudinal statistical approach that is robust to missing data, but note that these approaches are not general panaceas because they typically require assumptions about the randomness of missing data that are unrealistic, e.g., that the likelihood of missing outcome data is unrelated to any predictors (such as Time 1 severity).

INFERENCES: Are the Conclusions Supported by the Data? Are Cautions Noted When Appropriate?

Like problems that can occur with data analyses, problems with the inferences drawn from findings are too numerous to detail here. However, a common problem is the overly strong causal interpretation of correlational data. The limitations inherent in correlational, often retrospective data, need to be kept in mind when discussing even the most exciting results from an etiological model tested using structural equation modeling. Another common problem is overgeneralization of findings to nonstudied populations. Results from a treatment-seeking sample of Vietnam veterans with PTSD may not generalize to all veterans with PTSD, much less to all people with PTSD. Yet another problem is the clinical significance of findings, which is often confused with statistical significance. Authors are encouraged to report effect sizes for at least their major findings and to discuss the practical implications of their results.

FORMAT: Has JTS/APA Format Been Followed in the Reporting of all Statistical Information and in the Construction of Tables? Is the Number of Decimal Places Consistently Reported?

We ask that all authors consult the current (4th) edition of the *Publication Manual of the American Psychological Association* (APA, 1994) prior to submitting a manuscript. This volume provides necessary information about the required style used at the *Journal of Traumatic Stress* along with examples of correctly formatted tables and figures. It also provides useful writing tips, such as the correct tense to use to present findings in a results section (past) and to discuss findings in a discussion (present). Note that it also is necessary to consult the Instructions to Contributors on the inside back cover of the *Journal* for information about the *Journal's* points of deviation from APA style.

TABLES AND FIGURES: Are All Tables and Figures Necessary? (If Not, Suggest Deletions.) Are They Well Done?

Reviewers rarely ask for more tables and figures, and usually recommend that one or more tables or figures be deleted or combined with another. Several good rules of thumb are: avoid redundancy with the text; present in a table or figure only material that cannot be conveyed adequately in the text; make the table or figure interpretable without reference

to the text; and avoid unnecessarily busy or complicated tables and figures. For information about style, authors are asked to consult the *Publication Manual of the American Psychological Association* (APA, 1994) and the book by Tabachnick and Fidell (1996) for examples of tables prepared in APA format.

TERMINOLOGY: Are Technical Terms Used Correctly? Are All Statistical and Methodological Abbreviations Correct?

Authors are asked to consult the *Publication Manual of the American Psychological Association* (APA, 1994) as the best single source for information about technical terms and abbreviations. Perhaps the most commonly confused terms are "incidence" and "prevalence" (authors usually mean "prevalence"), but other terms are used incorrectly as well.

DETAILS: Are Numbers (e.g., Ns) Consistently Reported? Do Numbers (e.g., dfs) Appear to be Correct?

Statistical reviewers are asked to check the consistency of numbers reported. They also are asked to check for correctness insofar as possible. Readers are requested to review manuscripts prior to submission for these kinds of errors, which are especially likely in the case of resubmissions, e.g., an alteration made in response to a reviewer's concern that reduces the sample size is correctly reflected in the Results section but the Abstract and one table still mention the prior, larger, sample size.

Feedback to Authors

Statistical reviewers are asked to provide constructive criticism on these questions as needed. This feedback is sent to authors along with any content reviews as part of the Editor's response. In addition, statistical reviewers are asked to make a global assessment of a manuscript's quality and to rate the extent of revision needed, ranging from no revisions to significant revisions that cannot be corrected. These ratings are confidential and are not sent to authors. Instead, they are reviewed by the Statistical Editor along with the specific feedback that is given to authors, and a recommendation regarding necessary revisions is then made by the Statistical Editor to the Editor or Associate Editor who is handling a given manuscript.

Turnaround time—the amount of time between initial submission and editorial feedback—is a concern for authors and editorial staff alike. The statistical review process has been designed to minimize increases in turnaround and avoid long delays in providing feedback to authors. Statistical reviewers are given 2 weeks to conduct a review, compared with 1 month for content reviewers. Also, statistical reviewers are asked to refrain from conducting content review except insofar as necessary, usually with respect to the appropriateness of inferences permitted by a given design or statistical analysis. When possible, statistical reviews are returned by e-mail to further enhance timeliness of editorial response. We are satisfied with progress in this area so far and will refine procedures as necessary in order to continue providing authors with timely feedback.

Our experience with the process of statistical review during the past year indicates that the statistical reviewers helped to enhance the quality of many important and interesting articles. We thank the reviewers for the assistance they have provided thus far and hope that all readers find this new development at the *Journal of Traumatic Stress* to be a worthwhile change.

References

Aiken, L. M., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

American Psychological Association. (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington, DC: Author.

Borkovec, T. (1993). Between-group therapy outcome research: Design and methodology. *NIDA Research Monograph, 137,* 249-289.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cook, D. S., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis for field settings.* Chicago: Rand-McNally.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: Harper-Collins.