es exist or, as Guyll and Madon advocated, that findings for one ethnic group will generalize to another. I noted,

> The intent of this recommendation is not to imply that theories and principles lack generality. They may or may not. The point is that they require evidence, in the best tradition of science and the principle of skepticism, and cross-validation. (Sue, 1999, p. 1076)

I emphasized differences simply to point to the need for testing different populations and to stress that if psychologists have erred in their practice, it has been in the direction of ignoring other populations and cross-validation.

Finally, theoretical conservatism as defined by Guyll and Madon (2000) may be beneficial in certain ways. It provides a solid framework by which to judge research and prevents scientists from adopting theories based on whimsical fluctuations in research findings. On the other hand, theoretical conservatism may also discourage theoretical changes and valid counterpoints. (One should recall the devastating effects of theoretical conservatism centuries ago on challenges to the then prevailing thought that the sun rotated around the earth.) Yes, theoretical conservatism may be operating, but should not the scientific community aspire to a higher standard? Scientific conservatism, as opposed to Guyll and Madon's notion of theoretical conservatism, implies that one should not overgeneralize and should cross-validate findings. It is my belief that theories, models, and assessment instruments should all be rated according to adequacy with different populations. Those that have been used and found to be appropriate and applicable to many different populations should be considered as being more adequate, meeting more stringent requirements, and having greater goodness of fit to human beings than those that have been applied to only a few populations. Isn't this good science?

REFERENCES

Guyll, M., & Madon, S. (2000). Ethnicity research and theoretical conservatism. *American Psychologist, 55,* 1509–1510.

Sue, S. (1999). Science, ethnicity, and bias: Where have we gone wrong? *American Psychologist, 54,* 1070–1077.

Tebes, J. K. (2000). External validity and scientific psychology. *American Psychologist, 55,* 1508–1509.

Correspondence concerning this comment should be addressed to Stanley Sue, Department of Psychology, University of California, One Shields Avenue, Davis, CA 95616-8686. Electronic mail may be sent to ssue@ucdavis.edu.

DOI: 10.1037//0003-066X.55.12.1511

On the Documentation of Statistical Analyses in the "Clicky-Box" Era

Michael P. Kaschak and Colleen F. Moore
University of Wisconsin—Madison

The proliferation of user-friendly computer packages that perform statistical analyses has led to what we call the "clicky-box" era in statistics: It is now possible to run complicated statistical analyses by clicking on menu options and reading the output. Although the advent of such technology has had a profoundly positive impact on psychological research (think of doing a factor analysis or fitting a structural equation model without such software!), it has also led to a greater opportunity for confusion.

Errors in the handling and analysis of data are a perpetual problem in social science. Dawes (1988) cited an example of a publication submission withdrawn because of a statistical error. On careful reading, a published exchange in the *American Journal of Community Psychology* can be traced to a data entry error (Denney, 1988; Reinke, Holmes, & Denney, 1981). A PsycLIT search on the term *errata* showed that investigators do confess in print to analysis errors that alter the conclusions drawn in previously published work, as well as to other analytic errors (see, e.g., "Correction to Maylor," 1995; "Correction to McKinley and Nosofsky (1996)," 1998).

It is impossible to know whether undetected data analysis errors have increased or decreased with user-friendly statistical packages. However, a good number of statistical packages seem almost to invite errors to creep into the various stages of the data handling process by failing to record output information that is essential to the interpretation of the statistics: What data set was being analyzed? What subsample of observations within the data set was being analyzed? What transformations of the variables were performed to lead to this output?

Lacking documentation that links specific data collected to the statistical output presented in a publication is akin to lacking the chain of evidence in a criminal case. Establishing this chain of evidence is exceedingly important in science. Those psychologists who teach statistics frequently have graduate students visit them during office hours with a confusing jumble of pages of output from a user-friendly statistical program. The students often bring bits and pieces of several different analyses. Such students are usually analyzing data collected as part of a federally funded research program overseen by a professor. Typically, principal investigators on such programs are rather hands off in dealing with data in that they rely on statistical results produced by their graduate assistants and postdoctoral trainees. Finally, many reviewers of manuscripts and grant proposals find fundamental errors such as degrees of freedom that are inconsistent with either the reported sample size or the experimental design specified in the method section. Good documentation of data analyses can avoid such errors.

Some essential procedures for the documentation and safe handling of data follow. Required documentation includes:

1. Date analysis was run.
2. Name of data file analyzed.
3. Information about the data file:

   Date file was created and last altered;

   Whether or not the file contains a subset of a larger data file or represents data combined from several data sets;

   Whether or not the data in the file were screened for outliers, and if so, what criteria were used in the screening;

   Coding scheme for the data;

   Information about any transformations or subscores that were calculated on the data;

   What sample or population was used to generate the mean and standard deviation used in creating Z scores; and

   What factor analysis was the basis for a particular set of factor loadings.

Safety procedures in data analysis include:

1. Always print out descriptive statistics. Check the listed sample size, the range of scores, standard deviations, and degrees of freedom to be sure that nothing has gone wrong.
2. Make a new data file for each subset of data analyzed and for each major set of transformations performed on the data. Document the contents of each file.
3. Document what is in each data file in a notebook (electronic or hard copy). Make sure the file names are distinct enough that they will not be confused.
4. When transforming data, do a few hand calculations to make sure the program did what it was supposed to do.

These procedures take into account the iterative nature of statistical analysis as well as specific concerns that may arise with the analysis of particular types of data sets. This documentation is of special concern in large research labs, where analyses of a particular data set may be conducted by many different individuals or where many individuals collect different data that appear superficially similar. Clear documentation allows others who come in contact with the data sets and the output from the analyses to know exactly what was done and how it was done. In an age where data files can be transferred electronically, the possibility of data entry errors, miscodings of variables, and the like going undetected in large data sets is real.

It is easy to imagine how failure to accurately document a data set can lead to undesirable consequences. At the least, these consequences can be costly with regard to time, such as having to spend time figuring out what data were analyzed to produce the statistics on one's printouts (especially if one is returning to the printouts after some length of time, such as after receiving journal reviews) or having to repeat a set of analyses. More seriously, one could easily publish data that are incorrect in some way, and, in the worst case scenario, an accusation of academic fraud might result. Reese (1999) discussed how poor documentation of experimental protocol can cloud the literature with many failed attempts at replication. Some failures to replicate (either published or unpublished) could easily be due to differences in handling, or mishandling, of data in the original study or in the replication studies.

The procedures outlined above serve as a guide to help avoid the problems and confusions that can easily arise in the clicky-box era of highly automated data analysis. Of course, purveyors of clicky-box statistical packages could help all researchers by including some of the necessary information (date, data file used, and sequential page numbering) in the output files by default. Even with such a favorable change, however, the burden of establishing a chain of evidence from data collection to statistical results would still fall on the investigator.

REFERENCES

Correction to Maylor. (1995). *Psychology and Aging, 10,* 190. [Correction notice to Maylor, E. A. (1993). Aging and forgetting in prospective and retrospective memory tasks. *Psychology and Aging, 8,* 420–428.]

Correction to McKinley and Nosofsky (1996). (1998). *Journal of Experimental Psychology: Human Perception and Performance, 24,* 339. [Correction notice to McKinley, S. C., & Nosofsky, R. M. (1996). Selective attention and the formation of linear decision boundaries. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 294–317.]

Dawes, R. M. (1988). *Rational choice in an uncertain world.* New York: Harcourt Brace Jovanovich.

Denney, N. W. (1988). A reanalysis of the influence of a "friendly visitor" program. *American Journal of Community Psychology, 16,* 409–425.

Reese, H. W. (1999). Strategies for replication research exemplified by replications of the Istomina study. *Developmental Review, 19,* 1–30.

Reinke, B. J., Holmes, D. S., & Denney, N. W. (1981). Influence of a "friendly visitor" program on the cognitive functioning and morale of elderly persons. *American Journal of Community Psychology, 9,* 491–504.

Correspondence concerning this comment should be addressed to Michael P. Kaschak or Colleen F. Moore, Department of Psychology, University of Wisconsin—Madison, 1202 West Johnson Street, Madison, WI 53706. Electronic mail may be sent either to mpkaschak@students.wisc.edu or to cfmoore@facstaff.wisc.edu.