The Clinical Neuropsychologist, 2014
Vol. 28, No. 4, 556–574, http://dx.doi.org/10.1080/13854046.2012.762552

Routledge
Taylor & Francis Group

The STROBE Statement and Neuropsychology: Lighting the Way Toward Evidence-Based Practice

David W. Loring[1] and Stephen C. Bowden[2]

[1]Departments of Neurology and Pediatrics, Emory University School of Medicine, Atlanta, GA, USA

[2]Melbourne School of Psychological Sciences, University of Melbourne, Melbourne, Australia

Reporting appropriate research detail across clinical disciplines is often inconsistent or incomplete. Insufficient report detail reduces confidence in findings, makes study replication more difficult, and decreases the precision of data available for critical review including meta-analysis. In response to these concerns, cooperative attempts across multiple specialties have developed explicit research reporting standards to guide publication detail. These recommendations have been widely adopted by high impact medical journals, but have not yet been widely embraced by neuropsychology. The STROBE Statement (STrengthening the Reporting of Observational studies in Epidemiology) is particularly relevant to neuropsychology since clinical research is often based on non-funded studies of patient samples. In this paper we describe the STROBE Statement and demonstrate how STROBE criteria, applied to reporting of neuropsychological findings, will maintain neuropsychology's position as a leader in quantifying brain–behavior relationships. We also provide specific recommendations for data reporting and disclosure of perceived conflicts of interest that will further enhance reporting transparency for possible perceived sources of bias. In an era in which evidence-based practice assumes an increasingly prominent role, improved reporting standards will promote better patient care, assist in developing quality practice guidelines, and ensure that neuropsychology remains a vigorous discipline in the clinical neurosciences that consciously aspires to high methodological rigor.

Keywords: Professional issues; Statistical methods; STROBE statement; Observational studies; Critical appraisal.

INTRODUCTION

Clinical neuropsychology has a rich tradition in the scientist-practitioner model. Because of their research training during graduate education, neuropsychologists have the primary tools to advance the understanding of various disease-related effects on cognition and behavior through the publication of both funded and non-funded clinical research. The impact of many neuropsychological findings, unfortunately, is often diminished due to incomplete or inconsistent data acquisition and reporting detail. Maximizing reporting transparency and clearly identifying potential sources of bias not only provides explicit detail for readers to more easily evaluate a study's strengths and weaknesses, but also facilitates data synthesis across multiple research projects (e.g.,

Address correspondence to: David W. Loring, Ph.D., ABPP-CN, Department of Neurology, Emory University School of Medicine, 101 Woodruff Circle, Suite 6000, Atlanta, GA 30322, USA. Email: dloring@emory.edu

Accepted for publication: December 21, 2012. First published online: January 28, 2013.

systematic reviews or meta-analyses), which are increasingly used for evidence-based recommendations to inform clinical practice (Bilder, 2010; Chelune, 2010).

The influence of inadequate data analysis description in psychological studies has recently been highlighted (Simmons, Nelson, & Simonsohn, 2011). Incomplete data capture or data disclosure, however, is not limited to behavioral investigations. The National Institute of Neurological Disease and Stroke (NINDS), for example, addressed this concern by developing Common Data Elements (CDEs) for specific neurologic diseases and conditions (Adelson et al., 2012; Grinnon et al., 2012; Loring et al., 2011). Reporting standards have been addressed more broadly by the larger research community through the development of guidelines recommending specific information that should be included in research reports to ensure complete and accurate study detail disclosure. The STROBE Statement (STrengthening the Reporting of Observational studies in Epidemiology) was developed in response to the common problem of insufficient reporting detail necessary to fully establish a study's strengths and weaknesses in non-randomized designs, and offers specific recommendations to determine a study's generalizability (Vandenbroucke et al., 2007; von Elm et al., 2007).

Other reporting guidelines include the CONSORT Statement (CONsolidated Standards Of Reporting Trials; Begg et al., 1996; Schulz, Altman, Moher, & CONSORT Group, 2010), the STARD initiative (STAndards for Reporting of Diagnostic Accuracy; Bossuyt et al., 2003), and PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses; Liberati et al., 2009). Although checklists for reporting findings are primarily intended as publication guidelines, like NINDS CDEs, they identify important components of good clinical research that will ultimately improve the scientific value of the published literature. High-impact medical journals increasingly require use of appropriate reporting guidelines in their instructions to authors (Tao et al., 2011), and neurology journals such as *Neurology* and *The Lancet Neurology* require the presentation of clinical research according to appropriate checklists.

Like much of biomedical research, clinical neuropsychology research often relies on quasi-experimental designs or observational studies (Cook and Campbell, 1979) in which randomization cannot be conducted, and the STROBE Statement has special relevance. An observational study is broadly defined as one in which the etiology or diagnostic condition in a clinical population is pre-existing and is not under the control of the investigator. Consequently the investigator seeks to control the influence of alternative explanations for cause in order to make the strongest case possible for a relationship between clinical diagnosis and outcome variables. In the present article we briefly review the STROBE Statement in the context of neuropsychology and comment on its relevance to neuropsychology research reporting, although we also briefly discuss reporting standards for intervention trials as appropriate. We provide examples of how STROBE guidelines maximize transparency for multiple aspects of study conduct, and offer specific recommendations for data reporting and disclosure of perceived conflicts of interest that we consider to reflect important elements of "best practice" that are derived from current standards in statistics, research methodology, and clinical medicine. We hope that this paper will encourage the adoption of STROBE Statement criteria in the publication guidelines by appropriate neuropsychology editorial boards.

STROBE (STRENGTHENING THE REPORTING OF OBSERVATIONAL STUDIES IN EPIDEMIOLOGY)

The STROBE Statement (www.strobe-statement.org) was first published in 2007 to guide observational trial reporting with the goal of maximizing scientific transparency (Vandenbroucke et al., 2007; von Elm et al., 2007). Although a similar checklist for randomized controlled-trials (RCTs) or clinical-treatment studies was published in 1996 (Begg et al., 1996), a separate set of observational study guidelines was considered necessary because certain aspects of RCT design and reporting are not directly relevant for observational studies.

The STROBE Statement development committee included editorial staff from influential medical journals including *Annals of Internal Medicine, BMJ, Bulletin of the World Health Organization, International Journal of Epidemiology, JAMA, Preventive Medicine, and The Lancet*. In addition, prominent epidemiologists, methodologists, statisticians, and practitioners from Europe and North America also participated (von Elm et al., 2007). The development committee drafted the initial version of the STROBE Statement, and STROBE has been subsequently revised three times.

The STROBE checklist includes 22 content areas that are considered key elements for good research reporting (see Table 1). The checklist is not intended to be a rigid or fixed-formula reporting template, but rather to identify appropriate details for inclusion in research publications. Although adherence to the STROBE Statement does not pre-determine the quality or importance of observational research, increased transparency and reporting of study detail enhances the evaluation of study validity and helps identify design flaws or reporting omissions that might diminish a study's conclusions. The STROBE Statement includes a detailed exposition of its rationale and purpose, and we direct readers who wish to gain a more thorough understanding of the STROBE criteria to that article (Vandenbroucke et al., 2007). The rationale behind STROBE recommendations are clearly detailed and easily understood, and our comments reflect areas in which neuropsychology can increase precision to facilitate subsequent evidence-based characterization. Where appropriate, we also suggest specific neuropsychology reporting features to improve reporting detail. This commentary is intended to be read in conjunction with the STROBE checklist (Table 1).

COMMENTARY

STROBE Item 1: Title and Abstract

Good abstract detail that succinctly conveys essential study design information is increasingly important because a primary source of information for those without access to institutional electronic journal subscriptions is through public databases such as PubMed. Information from abstracts only will increase as smart phones and other smart devices become a primary link to the reported literature. Common terms that may be helpful in characterizing neuropsychology findings include "case-control study" or "retrospective review." A brief summary of the level of evidence provided by the study may also be included in the Abstract (see Oxford Centre for Evidence-Based Medicine: http://www.cebm.net/index.aspx?o=5653).

Table 1. STROBE Statement Checklist

Item	Item No	STROBE Recommendation
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract. (b) Provide in the abstract an informative and balanced summary of what was done and what was found.
Introduction		
Background/ rationale	2	Explain the scientific background and rationale for the investigation being reported.
Objectives	3	State specific objectives, including any pre-specified hypotheses.
Methods		
Study design	4	Present key elements of study design early in the paper.
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection.
Participants	6	*Cohort studies*: Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up. For matched studies, give matching criteria and number of exposed and unexposed participants. *Case-control studies*: Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls. For matched studies, give matching criteria and the number of controls per case. *Cross-sectional studies*. Give the eligibility criteria, and the sources and methods of selection of participants. *Cohort study*- For matched studies, give matching criteria and number of exposed and unexposed. *Case-control Study*- For matched studies, given matching criteria and the number of controls per case.
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.
Study size	10	Explain how the study size was arrived at.
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why.
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding. (b) Describe any methods used to examine subgroups and interactions. (c) Explain how missing data were addressed. (d) Cohort study—If applicable, explain how loss to follow-up was addressed. Case-control study—If applicable, explain how matching of cases and controls was addressed. Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy. (e) Describe any sensitivity analyses.
Results		
Participants	13	(a) Report numbers of individuals at each stage of study—e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analyzed. (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram.
Descriptive data	14	(a) Give characteristics of study participants (e.g., demographic, clinical, social) and information on exposures and potential confounders.

(*Continued*)

Table 1. (*Continued*)

Item	Item No	STROBE Recommendation
		(b) Indicate number of participants with missing data for each variable of interest.
		(c) Cohort study—Summarize follow-up time (e.g., average and total amount).
Outcome data	15	Cohort study—Report numbers of outcome events or summary measures over time.
		Case-control study—Report numbers in each exposure category, or summary measures of exposure.
		Cross-sectional study—Report numbers of outcome events or summary measures.
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included.
		(b) Report category boundaries when continuous variables were categorized.
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period.
Other analyses	17	Report other analyses done—e.g. analyses of subgroups and interactions, and sensitivity analyses.
Discussion		
Key results	18	Summarize key results with reference to study objectives.
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.
Generalizability	21	Discuss the generalizability (external validity) of the study results.
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based.

Reproduced with permission from von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). Strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. *BMJ, 335*(7624), 806–808. doi: 10.1136/bmj.39335.541782.AD

STROBE Items 2 and 3: Background/Rationale and Objectives

The Background/Rationale provides the general scientific and clinical justification for performing the study by discussing relevant literature, identifying gaps in knowledge, or highlighting areas of controversy that the research study is designed to address. Objectives are narrower, and they articulate specific research goals that, for example, focus on a clinical problem or question to be resolved, operational definitions of neuropsychological terms, and include explicit research hypotheses to be tested by the study.

A large number of variables comprise most clinical neuropsychology datasets, which gives rise to concern regarding Type I error rates when multiple neuropsychological variables are analyzed. Thus, in addition to the primary research hypothesis contained in Objectives, secondary and exploratory hypotheses can be described and which will decrease concerns over Type I error rate while balancing Type II error

risks. Secondary or exploratory hypotheses should be presented and clearly differentiated from the primary hypothesis(es) that form the initial basis for conducting the research (see also Simmons et al., 2011).

STROBE Item 4. Study Design

The study design must contain sufficient detail so that readers are able to independently judge the level of evidence contained in the report. The level of evidence provided by the study, whether or not reported in the Abstract, can also be justified in detail in the methods section (see Oxford Centre for Evidence-Based Medicine: http://www.cebm.net/index.aspx?o=5653). Important details for neuropsychological studies include whether the data were prospectively collected or retrospectively analyzed. In addition, full reporting of the data collection strategy for all variables is necessary to evaluate of the representativeness of reported results. For example, if data are obtained from clinical referrals rather than an explicit study protocol, how many patients had missing data that could not be analyzed? Common reasons for missing data include abbreviated assessment protocols for slower patients or from variations in test administration practice based on specific clinical referral questions. Thus it should be noted whether test administration and data collection varied by referral type.

STROBE Item 5: Setting

Data collection and patient contact setting are often implied rather than explicitly stated. Important setting characteristics include appropriate documentation of referral sources and referral questions and whether all patients in the same referral cohort are analyzed and reported. In some practices, for example, referrals for possible dementia include all patients with subjective memory deficit, whereas in others, dementia referrals are made based on low scores on cognitive screening tests such as the Montreal Cognitive Assessment (MoCA) (Nasreddine et al., 2012). Sources of clinical data affect base rates that may substantively influence important design and interpretive factors, and ultimately affect generalizability to clinical settings that may differ from the one in which the neuropsychological study data were obtained.

STROBE Item 6: Participants

The rationale for sample selection should be clear enough for readers to understand how to replicate the sample in all essential details (see also flow diagram use under STROBE Item 13 below). Matching participants on relevant variables should be described appropriately because matching on the primary variable of interest may create a systematic mismatch on other relevant dimensions (Meehl, 1975). For example, if disease effects on memory retrieval in Alzheimer disease (AD) are compared to those in multiple sclerosis (MS), the groups may be "matched" on cognitive impairment as reflected by the MoCA. However, matching on the MoCA creates a sample of AD patients with disproportionately preserved cognition as reflected by relatively high MoCA scores compared all AD patients, whereas the MS sample will have poor MoCA scores reflecting disproportionate cognitive impairment compared to all MS patients, decreasing the generalizability of results to each clinical group.

A common source of bias may result from matching study groups on "premorbid" IQ. In view of the difficulties in ascertaining estimates of premorbid IQ that work equally well in different clinical populations (Mathias, Bowden, Bigler, & Rosenfeld, 2007), unsuspected biases may arise for which there are no simple statistical controls. If matching is suspected to be a source of bias, both matched and unmatched comparisons should be performed and difference that may emerge can be identified and appropriately discussed. Lack of variation by matching method will enhance confidence in the generalizability of results.

STROBE Item 7: Variables

The method for establishing the clinical diagnosis should be presented in appropriate detail. Importantly for patient samples relying in part on behavioral features for diagnoses, were diagnoses fully independent of neuropsychological findings (i.e., is there any potential for criterion contamination of results due to confounding or inflated correlation between dependent and independent variables due to sample selection processes)? In epilepsy surgery series, for example, patients may be excluded as surgical candidates based on patterns of preoperative neuropsychological profiles suggesting unacceptable post-operative neuropsychological risk. Thus reported post-surgical cognitive outcome data cannot be used to establish risk of cognitive decline for all surgical candidates evaluated. This problem is akin to lacking the "untreated" or "diagnosis-absent" control statistics in a treatment or diagnostic validity study, and this may bias results. Although the epilepsy surgery example is obvious, biases in sampling strategy may not always be apparent but may give rise to covariance misestimation (Arkes, 1981; Meehl, 1975).

Defining groups using neuropsychological data introduces the risk of contamination bias between dependent and independent measures. For example, Mild Cognitive Impairment (MCI) is a behavioral diagnosis and is typically established based on neuropsychological memory test performance at or below the mean by some specified criterion. Studies of diagnostic accuracy or MCI clinical outcome often use the same neuropsychological memory test measure or other clinical variable as the dependent variable, although the same variable was employed as an independent variable to establish initial diagnosis. Study results, however, may be biased in the direction of supporting the validity of the dependent variable because it was also used as the independent variable to define groups.

A subtler form of contamination bias may arise if the same clinician knows the diagnostic status of patients and then administers or scores the test that is to be validated, even if different tests are used for initial diagnosis and outcome evaluation. Although many studies avoid these biases, the bias in outcome variable assessment is a particular risk in observational studies where there may be no formal mechanisms to blind the clinician involved in dependent variable assessment from knowledge of initial diagnosis. Retrospective or archival clinical data analysis is also particularly susceptible to these biases because the same team of clinicians has been involved in all steps in initial diagnosis and outcome assessment. Such biases reduce the likelihood of independent replication when the test, thought to be valid, is subsequently evaluated in a study that carefully controls these biases.

A slightly different issue is encountered when a clinical sample is restricted based on neuropsychological test scores (e.g., excluding patients with FSIQ <70) since this decreases the generalizability of findings. This is a problem when the mean ability level in the population of interest is reduced by (a) background factors such as educational or cultural opportunity or comorbid psychopathology, (b) primary disease effects such as TBI or seizure disorders, or (c) secondary disease effects such as depression. Reporting all patients meeting inclusion criteria without any exclusion or sample restructuring based on any neuropsychological test performance reduces this bias, and subsequent analyses performed upon matching variable stratification can be conducted with results reported both with and without matching. If there is a compelling reason to exclude patients based on some level of neuropsychological performance, then the number and percentage of excluded individuals should be reported, and the primary study results should be reported before and after exclusion. The analysis of secondary focus (e.g., before exclusion of a subset) can be typically reported very succinctly by stating that same pattern of results seen in the primary analysis was observed. Where differences exist between analyses before and after exclusion more careful consideration is required to defend the primary findings against threats of bias.

How the assessment data were obtained should be documented. Was the individual performing the neuropsychological testing a neuropsychologist, trainee, or paid technician, and did different individuals perform the assessments? If persons with different levels of responsibility and training performed the assessment, this should be disclosed and the total number of different individuals should be reported. Was there any quality control of data acquisition (e.g., double scoring, comparison of scoring interpretations, evaluation of inter-rater reliability) if multiple examiners were used? Clinically derived datasets should aspire to the same standard of quality control and reporting as funded research projects.

STROBE Item 8: Data Sources/Measurement

Psychometric characteristics of test scores employed in the study should be reported succinctly. At a minimum, this should include the reliability and convergent/discriminant validity evidence for the tests. In addition to the test manual the most recent independent test reviews should be cited including meta-analysis or systematic review whenever possible. Test psychometrics and the impact of reliability and validity on clinical evaluation of states and traits is a defining characteristic of applied psychological science, distinguishing neuropsychology from related disciplines such as behavioral neurology and neuropsychiatry (Adams, 1980; Einhorn, 1988), and this information should not be ignored when presenting research outcomes.

Although neuropsychology often reports normative or standardized scores, the size and quality of the normative samples across tests often differ, and the normative sampling techniques across tests may not be equivalent. If normative results are reported from tests with different normative samples, it may be necessary to highlight this difference. If transformed scores are derived from the study sample rather than calculated from independent normative datasets, non-transformed scores should also be presented. “Normalized” scores based on a study control group are often performed to facilitate the combination of performances across multiple tests into a single cognitive outcome (e.g., Altinbas et al., 2011). Reporting only normalized scores without indi-

vidual non-transformed individual test performances precludes individual test use in possible subsequent meta-analysis, however, and also prevents their clinical application given unknown individual test effect sizes described on an established clinical metric. The rationale for reporting a standardized score rather than raw score should be identified. Comparison of findings with standardized scores versus raw scores should be considered, and if discrepancies exist, should be reported and discussed. If no discrepancies exist, noting this in the results will increase confidence in study findings.

STROBE Item 9: Bias

Techniques to address sources of bias continue to be developed (Shrier and Platt, 2008). However, unique sources of bias in neuropsychology may include idiosyncratic scoring methods or diagnostic criteria that either differ based on specific tests used that may be similar but not equivalent (Loring et al., 2008) or that rely, in part, on judgment (e.g., various definitions of Mild Cognitive Impairment; Ritchie and Tuokko, 2010). Another important source of bias includes partial reporting of results, particularly when large numbers of variables are analyzed but only selected variables are reported (i.e., "fishing expedition"). Such selective reporting inflates the true Type I error rate, ultimately decreasing the likelihood of replication with an independent sample (Simmons et al., 2011). Attention to issues related to description of design, collected versus reported data, and analyzed versus unanalyzed results (as described in previous sections), will all aid to reduce bias.

STROBE Item 10: Study Size

How a sample size was determined often receives little formal comment in neuropsychology studies, and it typically is unclear if sample size was based on power estimates derived from the literature. Sample size and power are also relevant for interpretation of null-findings (Cohen, 1988). In line with the recommendations of the APA Task Force on Statistical Inference (Wilkinson, 1999), results should be reported in terms of effect size, confidence intervals around means, and conventional tests of statistical inference. This approach avoids reporting group comparisons as unimportant when interesting effects are present despite non-significant statistical test results, as may be the case with small sample studies of rare or unusual disease conditions.

STROBE Item 11: Quantitative Variables

If data transformations or recordings of raw scores were performed, what is their impact on reported findings? This is addressed, in part, in STROBE Item 8. It is common to reduce one or more continuous independent variables to a dichotomous classification system such as considering a memory test to be either "normal" or "impaired." However, dichotomization or other score categorization sacrifices significant information associated with between-person variance. Dichotomizing continuous variables also ignores the fact that most tests used in neuropsychology reflect arbitrary metrics in which any score reflects an indirect measure on the construct of interest such that the relationship between a one-unit change on the observed score and of change on the underlying dimension of disease impact may be unknown (Blanton &

Jaccard, 2006). Arbitrary dichotomization, however, may nevertheless imply good understanding of disease characterization. Comments should be included to address whether the same pattern of results is present when data are analyzed in a way that preserves the quantitative parametric information of the variables of interest (e.g., multiple regression analysis).

STROBE Item 12: Statistical Methods

Although APA format combines description of statistical methods with reporting of results (covered under STROBE Items 13–16 below), justification of analytic methods should not be neglected and whether the assumptions underlying any statistical test are satisfied should be evaluated and reported (see Gravetter & Wallnau, 2007; Howell, 1992; Tabachnick & Fidell, 2007). If test assumptions are satisfied, then a single sentence stating that the assumptions of the statistical test were met should be sufficient. If test assumptions are not met, the violation should be identified and methods to address the violation should be presented that typically include an alternative analytic approach (including evaluation of relevant assumptions for the alternative approach). If a statistical consultant is engaged to assist with data analysis, the consultant should be familiar with the APA guidelines on presenting results including test assumptions (Wilkinson, 1999) as well as recent trends in good practice (e.g., Cummings & Finch, 2005) in order that the consultant can provide relevant reporting information to the study authors for inclusion.

An approach to examining statistical test assumptions can be illustrated using correlational analysis. The Pearson correlation coefficient assumes a linear relationship between two variables as well as normality in the distribution of the two variables being compared, but this assumption is rarely evaluated. When the assumption of normality is not met, the magnitude of the correlation may be substantially inflated or attenuated by one or more outliers. The assumptions of linearity and normality can be addressed by examining the bivariate scatter plot and distributional statistics and stating there was no obvious evidence of non-linearity between the respective variables or departures from normality. When the normality assumption is not met, the non-parametric Spearman correlation that does not require distribution normality can be used instead.

Additional detail of the approaches used to address study design methods (e.g., matching samples on background variables) should be presented. Methods to address sampling mismatch should also be included (e.g., analysis of covariance), and if statistical methods or re-sampling are used to address sampling mismatch, adjusted and unadjusted results should be reported to allow the reader to assess their effect.

Power and confidence intervals have been addressed above (STROBE Item 10). Outcome statistics, however, should include sensitivity and specificity analysis for a diagnostic study, or risk reduction or numbers needed to treat should be reported for an intervention study. These clinical impact statistics should also always be reported with confidence interval boundaries (Straus, Richardson, Glasziou, & Haynes, 2010), not only to communicate the likely range of the observed statistic, but also the precision of estimate. Wider confidence intervals, typically observed with smaller samples, convey less precision.

STROBE Item 13: Participants

Inclusion and exclusion criteria should be reported in appropriate detail, including whether the study sample was derived from a larger pool or clinical database. Patient groups that are regularly referred for evaluation are likely more representative of the entire population compared to patients who are infrequently sent for testing based on an unusual presentation or characteristic. The source of patient referral and percentages (e.g., general medicine, psychiatry, plaintiff versus defense attorney) and percentages of total referrals should also be described. If participants are excluded from analysis due to failure to meet specified inclusion criteria, their numbers should be identified. Patients not meeting entry criteria are called "screen failures" in clinical trials, and presenting information on screen failures reduces the problem of unrecognized or inadvertent differences on important variables (e.g., sex or ethnic group) that may limit generalizability. Discursive description of these sometimes multiple selection processes and missing data patterns are detailed in the STROBE Statement (Vandenbroucke et al., 2007). A flow diagram is often very helpful to clarify participant recruitment, selection, and transition through the data collection phases. Although demanding of journal space, a flow diagram is easily incorporated into a journal's online components associated with its published content.

STROBE Item 14: Descriptive Data

Neuropsychology reports often meet this expectation described in STROBE but quality of data reporting varies. Descriptive statistics for demographic and outcome variables is critical to establish sample characteristics and representativeness. Means, standard deviations, and range for interval scales; median and inter-quartile range for ratings and ordinal scales; and frequencies for nominal scales for all relevant variables should be reported routinely, together with details of departures from normality as necessary.

STROBE Item 15: Outcome Data

In line with APA publication guidelines (Wilkinson, 1999), outcome data and results are typically presented together in the same section. Consequently, STROBE Items 15 and 16 should be considered as part of the same recommendations. Appropriate descriptive statistics including means and standard deviations (or clearly labeled standard errors of the mean) with respective sample sizes and 95% confidence intervals for sample means are desirable for all study variables and for each sample or cohort in order to facilitate subsequent meta-analyses. Use of figures alone without this degree of data precision, unless these data are available as a web resource, should be avoided.

STROBE Item 16: Main Results

Reporting of study findings is a strength of many neuropsychological studies, and psychologists are well trained in this aspect of research. However, there are often multiple approaches to reporting findings that can be considered, and excellent guides are available for reporting of study results both in terms of statistical significance and

confidence intervals, but also in contemporary effect-size or "clinical importance" metrics (Cumming and Finch, 2005; Straus, Richardson, Glasziou, & Haynes, 2010; Wilkinson, 1999). Guidelines for reporting statistical copy in psychology journals are also readily available (APA Publication Manual, 2009; Wilkinson, 1999). The importance of clearly describing analytic methods and justifying statistical tests, including tests for assumptions underlying each test, is highlighted above in STROBE Item 12. Inadequate statistical copy is a common source of ambiguity in published studies when results are scrutinized for clinical importance using critical appraisal techniques (Straus et al., 2010). In line with ethical guidlelines of major psychological associations, authors should make data available to colleagues who wish to verify results.

Detailed discussion of research design and analysis is beyond the scope of this article, although several points are worth mentioning. Clinicians and clinical researchers require skills in study design and analysis to fashion answerable clinical questions and to promote clinically relevant, rigorous research (Bowden, Harrison, & Loring, 2014). It is important that, if clinincal researchers work with statistical consultants, the statistical "tail" must not wag the clinical-conceptual "dog."

An important advance in statistical reporting over the last decade involves a stronger emphasis on practical impact of research findings rather than enslavement to null-hypothesis significance testing and the tyrany of $p < .05$ (Cohen, 1994). One manifestation of this trend is evident in the strong emphasis on reporting confidence intervals associated with inferential statistics (Cumming and Finch, 2005; Wilkinson, 1999). Confidence intervals also provide a direct means by which to estimate the precision of estimation of population parameters (e.g., group means). As with estimates of test scores in individual assessment, group mean estimates that are imprecise and have wide confidence intervals reflect a greater range of uncertainty over whether a mean value may vary if sampled on another occasion. Displaying the confidence interval for an individual's score or for a sample mean is an effective technique to communicate the precision in psychological trait estimation.

The focus on practical impact extends beyond statistical inference. If statistically significant results are observed for a diagnostic validity study or treatment intervention, then the results should also be presented in terms of effect size and clinical importance. Most neuropsychologists are familiar with reporting statistical findings in terms of effect size, and the choice of effect measure depends on the particulars of the analysis (Cohen, 1988; Wilkinson, 1999). Researchers should report one or more of the following as appropriate: (1) Cohen's d for differences between pairs of means, (2) R^2 for explained variance attributable to each independent variable in regression modelling, (3) Cohen's w for contingency table analysis using χ^2, or (4) partial eta-squared for variance attributable to independent variables or interaction terms in analysis of variance. Alternative measures of effect magnitude may be appropriate for other types of analysis (Cohen, 1988).

The other major focus of pratical impact involves reporting classification accuracy for diagnostic validity studies or treatment benefit in intervention studies. For diagnostic validity studies, sensitivity and specificity reporting is increasingly common, but this should be routine and include confidence intervals for all sensitivity and specificity estimates (Bowden & Loring, 2009; Straus et al., 2010). Strong clinical inferences continue to be drawn from reports presenting sensitivity and specifiity classification metrics alone, although those inferences would often change if the

confidence intervals associated with the point estimates were presented and appropriately considered. Authors should report probability revision curves with their sensitivity and specificity analyses (Frederick & Bowden, 2009; Straus et al., 2010) which illustrate the positive predictve power (PPP) and negative predictive power (NPP) of tests across the full range of prevalence from zero to one. Use of probability revision curves decrease the likelihood that clinicians will rely on fixed PPP or NPP valules derived from a single study that would not apply to other clinical settings with different base-rates (Arkes, 1981; Baldessarini, Finklestein, & Arana, 1983; Faust, 2003; Woods, Weinborn, & Lovejoy, 2003). Sensitivity and specificity values can also be converted to likelihood ratios and ilustrated with a likelhood ratio nomogram, which faciltates conversion of pre-test probability to post-test probability across the prevalence range (see Bowden & Loring, 2009; Grimes & Schulz, 2005).

For treatment interventions, practical impact reporting now extends well beyond reporting the test of significance on the post-treatment comparison between the treated versus the control group. Treatment effects are best reported as relative and absolute risk reduction together with the estimate of number needed to treat, all with respective confidence intervals. Absolute risk reduction and relative risk reduction should both be reported since, with low-prevance conditions, relative risk reduction may convey the impression of an important treatment that has small absolute benefit. Odds ratios are a common treatment effect metric but again, if reported, should be accompanied by the number needed to treat estimate since the latter are more intutitively accessible to the lay consumer as well as professionals (for detailed descriptions see Straus et al., 2010).

STROBE Item 17: Other analyses

See comments under STROBE Items 13–16 above.

STROBE Item 18: Key Results

This is a necessary component of good research reporting, relating the important findings directly back to study objectives that provided the scientific motivation to conduct the study. Both positive and null results should be described, along with unexpected results derived from secondary or exploratory analyses (Simmons et al., 2011).

STROBE Item 19: Limitations

Authors should address the role of referral source or clinician's practice style as possible bias, particularly if results are markedly different from reports in the published literature. Limitations may also include non-consecutive series in retrospective data collection (e.g., chart review) and any other important consideration that limits the generalizability of the results.

STROBE Item 20: Interpretation

Caution is emphasized by STROBE. Although multiple analyses do not negate specific findings, it is often times the patterns of scores across multiple neuropsychological variables that provides the most compelling evidence. Although

Type I error rate is often addressed by control of multiple analysis (but see Simmons et al., 2011), Type II error risk is less frequently addressed, which is of great potential concern should genuine group differences or differential risks be overlooked. Finally there is no universally agreed approach to managing Type I error rate, or even what is the best conceptual unit to calculate the error rate. One "food for thought" argument over the conceptual unit of error rate suggests that a researcher who is very productive and conducts many independent experients might consider a more conservative statisticial threshold compared to researchers who perform only a few independent studies over their career (i.e., conceptual unit for error rate is "career - wise" rather than "family-wise" or"experiment-wise"). When muliple papers are derived from the same (or overlapping) clinical dataset, authors should be explicit when describing research design elements that inflate the Type I error rate, and consider employing a more conservative critical p - value to infer statistical significance (Simmons et al., 2011).

STROBE Item 21: Generalizability

The generalizability discussion can take many different forms, but should seek to provide a balanced, realistic evaluation of the external validity of the study findings. Discussion focus may include the degree to which a specific magnitude of neuropsychological effect may be related to a limitation in real life behavior akin to Meehl's distinction between statistical and clinical significance (Meehl, 1973). Generalizability discussion may also include issues of referral source bias or differences in base rates previously discussed, and consideration of the precision of results obtained from report of confidence intervals around descriptive or inferential statistics. In this context the confidence intervals provide an estimate of the likely range over which estimates of critical results might fall if the study were to be replicated, together with an indication of the precision of the reported results (Cumming & Finch, 2005).

STROBE Item 22: Funding

Formal funding from granting agencies is straightforward. What should also be disclosed are other financial relationships and direct or indirect salary/income support. The most obvious examples for neuropsychology include to what degree the investigator's income is based on clinical evaluations, and is there is any financial stake in tests or measures involved in the report that have conflict of interest potential. The journal *Neurology*, published by the American Academy of Neurology, requires extensive disclosures across a range activities to ensure that readers are aware of possible source of bias from the authors that exend beyond monetary ones since there are considereable non-financial factors that may bias an investigator (Meador, 2009). *Neurology*'s disclosure list was developed to ensure maximum transparency and includes service on scientific advisory boards or serving on journal editorial boards (Knopman, Baskin, Pieper, Quimby, & Gross, 2011). Additional disclosures are made for gifts, funding for travel, patents held or pending, royalties from publishing, honoraria, corporate appointments and consultancies, memberships in speakers' bureaus, income derived from clinical or imaging studies in practice, research support (from commercial, government, academic, and foundations sources), and financial compensation from stock options,

license fee payments, royalties from technology, and legal procedings (expert testimony or affidavits).

Although it may not be appropriate for all neuropsychology studies, characterization of referral patterns in reports that have relevance for medico-legal or forensic evaluations should be reported. The percentage of plaintiff or defense referrals may be relevant both in relative and absolute terms, and as suggested by *Neurology*'s disclosure requirements, the percentage of income derived from expert witness fees and testimony (not fees generated from clinical evaluation) may be appropriate for the reader to know. Current medical disclosure convention is to detail all individual relationships in which at least $10,000 in income is generated in any given year. Similary, for investigators with a financial stake in assessment materials being reported, that relationship should be disclosed and the magnitude of that relationship described, irrespective of the dollar value, and whether direct or in kind. There is not only an extensive literature on potential monetary and secondary-gain influences on performance and clinical outcomes (Binder & Rohling, 1996; Harris, Mulford, Solomon, van Gelder, & Young, 2005), but also an emerging literature on financial conflicts of interest and how they influence research reporting (Bekelman, Li, & Gross, 2003; Okike, Kocher, Mehlman, & Bhandari, 2008; Perlis, Harwood, & Perlis, 2005). It is illogical to assume that patients are sometimes susceptible to the influences of financial gain or other incentives but that the professional community is immune to such effects. Neuropsychology journal editors should carefully consider comprehensive disclosure requriements to ensure that every article published is "free of bias, spin, and secondary gain" (Knopman et al., 2011).

DISCUSSION

Many neuropsychological publications are based on samples of convenience reflecting performances of patients who have been referred for neuropsychological evaluation to address specific clinical questions. Consequently group composition, at least in part, may be beyond the control of the experimenter (e.g., participants with specific clinical diagnoses or involve referral bias; Cook & Campbell, 1976). Relying on samples of convenience increases the need to maximize transparency when reporting neuropsychological findings because the risks of unacknowledged or unrecognized bias are greater than in studies where strict randomization can be used to control (though not necessarily eliminate) bias and confounding interpretations. The STROBE Statement offers explicit guidance to characterize sample composition and subsequent analysis to facilitate the identification of biased or unrepresentative results. Because sample proportions in observational studies rely on participant variables such as diagnoses, age, financial status, and many other variables that cannot be randomized, observational studies carry significant risks of selection and analytic bias that need to be properly documented so that results can be interpreted appropriately.

A detailed explanation and elaboration article discussing each STROBE checklist item and giving methodological background and published examples of transparent reporting in the context of biomedical research is freely available on several web sites:

PLoS Medicine at http://www.plosmedicine.org/
Annals of Internal Medicine at http://www.annals.org/
Epidemiology at http://www.epidem.com/

Additional information on the STROBE initiative together with regular updates is available at www.strobe-state. The STROBE Statement is referred to in the Uniform Requirements for Manuscripts Submitted to Biomedical Journals by the International Committee of Medical Journal Editors.

Greater reporting transparency enhances the interpretation of validity and benefits the scientific status of neuropsychology. With supplementary web-based tables and figures increasingly used in scientific publication, formal print page limitations no longer exert the same restriction on the amount of material that can be presented with a research report as in the past. The need for proper sample documentation is well known by neuropsychology researchers. However, the issues addressed by the STROBE Statement provide a convenient and accessible account of good clinical design and reporting that are recognized by neuropsychology as having potential influences on research outcomes (Bilder, 2010; Chelune, 2010). In addition, the method of *Critically Appraised Topics* (see Straus et al., 2010; www.cebmh.com; Bowden et al., in press) provides a structured step-by-step guide for clinicians to incorporate review of study methods and quality into clinical thinking and decision making. By increasing focus on explicit criteria by which to evaluate study methods such as the STROBE criteria (or more broadly to include STARD for diagnostic accuracy studies and PRISMA for meta-analyses), clinicians, researchers, and the journal publication process in neuropsychology will benefit, facilitating the adoption of comparable rigor as that required in high-impact medical journals.

ACKNOWLEDGMENT

We thank Myriam Cevallos, the anonymous reviewers from the STROBE group, and our anonymous *TCN* reviewers for their helpful comments on earlier drafts of this manuscript.

REFERENCES

Adams, K. M. (1980). In search of Luria's battery: A false start. *Journal of Consulting and Clinical Psychology, 48*, 511–516. doi: 10.1037/0022-006X.48.4.511

Adelson, P. D., Pineda, J., Bell, M. J., Abend, N. S., Berger, R. P., Giza, C. C., … Wainwright, M. S. (2012). Common data elements for pediatric traumatic brain injury: Recommendations from the working group on demographics and clinical assessment. *Journal of Neurotrauma, 29*, 639–653. doi: 10.1089/neu.2011.1952

Altinbas, A., van Zandvoort, M. J. E., van den Berg, E., Jongen, L. M., Algra, A., Moll, F. L., … van der Worp, H. B. (2011). Cognition after carotid endarterectomy or stenting. *Neurology, 77*, 1084–1090. doi: 10.1212/WNL.0b013e31822e55b9

American Psychological Association. (2009). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

Arkes, H. R. (1981). Impediments to accurate clinical judgment and possible ways to minimize their impact. *Journal of Consulting and Clinical Psychology, 49*, 323–330. doi: 10.1037/0022-006X.49.3.323

Baldessarini, R. J., Finklestein, S., & Arana, G. W. (1983). The predictive power of diagnostic tests and the effect of prevalence of illness. *Archives of General Psychiatry, 40*, 569–573. doi: 10.1001/archpsyc.1983.01790050095011

Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., … Stroup, D. F. (1996). Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA, 276*, 637–639. doi: 10.1001/jama.1996.03540080059030

Bekelman, J.E., Li, Y., & Gross, C. P. (2003). Scope and impact of financial conflicts of interest in biomedical Research. *JAMA, 289*, 454–465. doi: 10.1001/jama.289.4.454

Bilder, R. M. (2010). Neuropsychology 3.0: Evidence-based science and practice. *Journal of the International Neuropsychological Society, 17*, 1–7. doi: 10.1017/S1355617710001396

Binder, L. M., & Rohling, M. L. (1996). Money matters: A meta-analytic review of the effects of financial incentives on recovery after closed-head injury. *American Journal of Psychiatry, 153*, 7–10.

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist, 61*, 27–41. doi: 10.1037/0003-066X.61.1.27

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., … de Vet, H. C. (2003). Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *BMJ, 326*, 41–44. doi: 10.1136/bmj.326.7379.41

Bowden, S. C., Harrison, E. J., & Loring, D. W. (2014) Evaluating Research for Clinical Significance: Using Critically Appraised Topics to Enhance Evidence-based Neuropsychology. *The Clinical Neuropsychologist*, 28 (4), 653–668.

Bowden, S. C., & Loring, D. W. (2009). The diagnostic utility of multiple-level likelihood ratios. *Journal of the International Neuropsychological Society, 15*, 769–776. doi: 10.1017/S1355617709990373

Chelune, G. J. (2010). Evidence-based research and practice in clinical neuropsychology. *The Clinical Neuropsychologist, 24*, 454–467. doi: 10.1080/13854040802360574

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Cohen, J. (1994). The earth is round (p <.05). *American Psychologist, 49*, 997–1003. doi: 10.1037/0003-066X.49.12.997.

Cook, T. D., & Campbell, D. T. (1976). The design and conduct of true experiments and quasi-experiments in field settings. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 223–226). Skokie, IL: Rand McNally.

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

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist, 60*, 170–180. doi: 10.1037/0003-066X.60.2.170

Einhorn, H.J. (1988). *Accepting error to make less error*. New York: Cambridge University Press.

Faust, D. (2003). Holistic thinking is not the whole story. *Assessment, 10*, 428–441. doi: 10.1177/1073191103259534

Frederick, R. I., & Bowden, S. C. (2009). The test validation summary. *Assessment, 16*, 215–236. doi: 10.1177/1073191108325005.

Gravetter, F. J., & Wallnau, L. B. (2007). *Statistics for the behavioral sciences* (7th ed.). Belmont, CA: Thomas Wadsworth.

Grimes, D. A., & Schulz, K. F. (2005). Refining clinical diagnosis with likelihood ratios. *Lancet, 365*, 1500–1505. doi: 10.1016/S0140-6736(05)66422-7

Grinnon, S. T., Miller, K., Marler, J. R., Lu, Y., Stout, A., Odenkirchen, J., … Kunitz, S. (2012). National Institute of Neurological Disorders and Stroke Common Data Element Project: Approach and methods. *Clinical Trials, 9*, 332–339. doi: 10.1177/1740774512438980

Harris, I., Mulford, J., Solomon, M., van Gelder, J. M., & Young, J. (2005). Association between compensation status and outcome after surgery. *JAMA, 293*, 1644–1652. doi: 10.1001/jama.293.13.1644

Howell, D. C. (1992). *Statistical methods for psychology* (4th ed.). Boston, MA: PWS-Kent Publishing.

Knopman, D. S., Baskin, P. K., Pieper, K. M., Quimby, S. L., & Gross, R. A. (2011). Reporting potential bias. *Neurology, 76*, 110–112. doi: 10.1212/WNL.0b013e318206f646

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Ioannidis, J. P. … Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ*, 339, b2700. doi: 10.1136/bmj.b2700

Loring, D. W., Lowenstein, D. H., Barbaro, N. M., Fureman, B. E., Odenkirchen, J., Jacobs, M. P., & Stout, A. (2011). Common data elements in epilepsy research: Development and implementation of the NINDS epilepsy CDE project. *Epilepsia, 52*, 1186–1191. doi: 10.1111/j.1528-1167.2011.03018.x

Loring, D. W., Strauss, E., Hermann, B. P., Barr, W. B., Perrine, K., Trenerry, M.R., … Bowden, S. C. (2008). Differential neuropsychological test sensitivity to left temporal lobe epilepsy. *Journal of the International Neuropsychological Society, 14*, 394–400. doi: 10.1017/S1355617708080582

Mathias, J. L., Bowden, S. C., Bigler, E. D., & Rosenfeld, J. V. (2007). Is performance on the Wechsler Test of Adult Reading affected by traumatic brain injury? *British Journal of Clinical Psychology, 46*, 457–466. doi: 10.1348/014466507X190197

Meador, K. J. (2009). Re: Conflicts of interest for authors of American Academy of Neurology clinical practice guidelines. *Neurology, 72*, 1283. doi: 10.1212/01.wnl.000034632371448.85

Meehl, P. E. (1973). *Psychodiagnosis: Collected papers*. Minneapolis, MN: University of Minnesota Press.

Meehl, P. E. (1975). Nuisance variables and the ex post facto design. In M. Radner & S. Winokur (Eds.), *Analyses of theories and methods of physics and psychology* (pp. 373–402). Minneapolis, MN: University of Minnesota Press.

Nasreddine, Z. S., Phillips, N., Chertkow, H., Rossetti, H., Lacritz, L., Cullum, M., et al. (2012). Normative data for the Montreal Cognitive Assessment (MoCA) in a population-based sample. *Neurology, 78*, 765–766. doi: 10.1212/01.wnl.0000413072.54070.a3

Okike, K., Kocher, M. S., Mehlman, C. T., & Bhandari, M. (2008). Industry-sponsored research. *Injury, 39*, 666–680. doi: 10.1016/j.injury.2008.02.013

Perlis, C. S., Harwood, M., & Perlis, R. H. (2005). Extent and impact of industry sponsorship conflicts of interest in dermatology research. *Journal of the American Academy of Dermatology, 52*, 967–971. doi: 10.1016/j.jaad.2005.01.020

Ritchie, L. J., & Tuokko, H. (2010). Patterns of cognitive decline, conversion rates, and predictive validity for 3 models of MCI. *American Journal of Alzheimer's Disease and Other Dementias, 25*, 592–603. doi: 10.1177/1533317510382286

Schulz, K. F., Altman, D. G., Moher, D., & CONSORT Group (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *BMJ, 340*, c332. doi: 10.1136/bmj.c332

Shrier, I., & Platt, R. (2008). Reducing bias through directed acyclic graphs. *BMC Medical Research Methodology, 8*, 70–. doi: 10.1186/1471-2288-8-70

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology. *Psychological Science, 22*, 1359–1366. doi: 10.1177/0956797611417632

Straus, S. E., Richardson, W. S., Glasziou, P., & Haynes, R. B. (2010). *Evidence-based medicine: How to practice and teach EBM* (4th ed.). Edinburgh, UK: Elsevier Churchill-Livingstone.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Boston, MA: Allyn & Bacon/Pearson Education.

Tao, K -M., Li, X -Q., Zhou, Q. -H., Moher, D., Ling, C. -Q., & Yu, W. -F. (2011). From QUOROM to PRISMA: A survey of high-impact medical journals' Instructions to authors and a review of systematic reviews in anesthesia literature. *PLoS ONE, 6*, e27611. doi: 10.1371/journal.pone.0027611

Vandenbroucke, J. P., von Elm, E., Altman, D. G., Gøtzsche, P. C., Mulrow, C. D., Pocock, S. J., … Egger, M. (2007). Strengthening the reporting of observational studies in epidemiology (STROBE): Explanation and elaboration. *Epidemiology, 18*, 805–835. doi: 10.1097/EDE.1090b1013e3181577511

von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). Strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. *BMJ, 335*, 806–808. doi: 10.1136/bmj.39335.541782.AD

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594–604. doi: 10.1037/0003-066X.54.8.594

Woods, S. P., Weinborn, M., & Lovejoy, D. W. (2003). Are classification accuracy statistics underused in neuropsychological research? *Journal of Clinical and Experimental Neuropsychology, 25*, 431–439. doi: 10.1076/jcen.25.3.431.13800