JOURNAL OF PERSONALITY ASSESSMENT, *81*(3), 191–193

GUIDELINES EDITORIAL

Guidelines for Reporting Information in Studies of Diagnostic Test Accuracy: The STARD Initiative

Gregory J. Meyer
Department of Psychology
University of Toledo

Lijmer et al. (1999) recently examined the impact of methodological design features on the magnitude of validity coefficients obtained in studies examining the diagnostic accuracy of tests. From 11 meta-analyses, they examined the validity of 18 different medical tests across 218 data samples and found clear evidence that certain methodological qualities produced biased and overly rosy impressions about test validity. In general, studies that had lower methodological quality overestimated the diagnostic performance of a test. Some of the most important methodological artifacts emerging in their analyses were the use of a nonrepresentative group of patients (i.e., comparing a sample of designated cases to matched controls rather than systematically sampling within a representative clinical setting), the use of one set of procedures to verify positive test results but a different set of procedures to verify negative test results (e.g., surgical verification for patients with positive test results but verification by extended observation for those with negative test results), inadequate descriptions of the test procedures or population under study, and lack of sufficient blinding between the test predictor and criterion measure (e.g., criterion contamination).

Subsequently, the Cochrane Diagnostic and Screening Test Methods Working Group established a committee charged with improving the way information is reported in diagnostic studies as an initial step toward improving the quality of published research. A panel of 25 experts was convened to develop the Standards for Reporting of Diagnostic Accuracy (STARD), and early this year their initial report was published simultaneously in a number of prominent medical journals including *Journal of the American Medical Association, British Medical Journal, Lancet, Clinical Chemistry, Annals of Internal Medicine, Radiology, American Journal of Clinical Pathology, Clinical Biochemistry,* and *Clinical Chemistry and Laboratory Medicine.* The documents are freely available without copyright restrictions and can be obtained on the *Journal of Personality Assessment* (*JPA*) Web page under the "Resources for Reviewers and Authors" section (www.erlbaum.com/jpainfo.htm).

The STARD initiative provides two key guides that should make it easier for researchers, reviewers, and journal readers to evaluate the methodological quality of a diagnostic study and determine the likelihood for bias to be present in the findings. The first guide is a 25-item checklist, which is reproduced in Table 1 (with slight modification). The second guide is a prototypical flow diagram (see Figure 1) that provides detailed information about the number of patients present and classified at each stage of the study. The value of the flowchart is that it provides transparent information about the design of the research, including the method for recruiting patients, the order of testing, and the number of patients who undergo the index test and the criterion evaluation.

As of this issue of *JPA*, researchers submitting studies examining the diagnostic accuracy of tests are strongly encouraged to adhere to the STARD guidelines. This can be most readily accomplished by submitting a manuscript accompanied by a completed copy of the STARD checklist and by including in the manuscript a study-specific flowchart that visually reveals the procedures used to sample patients and obtain data from the predictor test and criterion. Templates for both the STARD checklist and flowchart are available on the *JPA* Web page. In addition, *JPA* reviewers are strongly encouraged to rely on the STARD guidelines when evaluating the suitability of a manuscript for publication.

It should be appreciated that much of the information recommended in the STARD checklist and flowchart would be valuable to report in most studies examining the validity of personality assessment instruments, not just those examining the test accuracy for diagnostic classification purposes. Thus, all researchers who anticipate submitting a manuscript to *JPA* are encouraged to review the STARD guidelines.

To help initiate the new *JPA* policy with respect to the STARD guidelines, the Statistical Developments and Applications section in this issue contains a very practical and

TABLE 1
The STARD Checklist for Reporting Information in Diagnostic Accuracy Studies

Section and Topic	*Item*	*Description*	*On Page No.*
Title, abstract, and keywords	1	Identify the article as a study of diagnostic accuracy (recommend keyword for PsycINFO "diagnostic efficiency"; recommended MeSH heading for Medline "sensitivity and specificity")	
Introduction	2	State the research questions or aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups	
Methods:			
Participants	3	Describe the study population: the inclusion and exclusion criteria and the settings and locations where the data were collected	
	4	Describe participant recruitment: Was this based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	
	5	Describe participant sampling: Was this a consecutive series of participants defined by selection criteria in items 3 and 4?; if not, specify how participants were further selected	
	6	Describe data collection: Was data collection planned before the index tests and reference standard were performed (prospective study) or after (retrospective study)?	
Test methods	7	Describe the reference standard and its rationale	
	8	Describe technical specifications of material and methods involved, including how and when measurements were taken, or cite references for index tests or the reference standard, or both	
	9	Describe definition of and rationale for the units, cut-off points, or categories of the results of the index tests and the reference standard	
	10	Describe the number, training, and expertise of the persons executing and reading the index tests and the reference standard	
	11	Were the readers of the index tests and the reference standard blind (masked) to the results of the other test?; describe any other clinical information available to the readers of the index test	
Statistical methods	12	Describe methods for calculating or comparing measures of diagnostic accuracy and the statistical methods used to quantify uncertainty (e.g., 95% confidence intervals)	
	13	Describe methods for calculating test reproducibility (e.g., interrater reliability) if done	
Results:			
Participants	14	Report when the study was done, including beginning and ending dates of recruitment	
	15	Report clinical and demographic characteristics (e.g., age, sex, spectrum of presenting symptoms, comorbidity, current treatments, and recruitment center)	
	16	Report how many participants satisfying the criteria for inclusion did or did not undergo the index tests or the reference standard, or both; describe why participants failed to receive either test (a flow diagram is strongly recommended)	
Test results	17	Report the time interval from index tests to reference standard, and any treatment administered between	
	18	Report the distribution of severity of disease (define criteria) in those with the target condition and other diagnoses in participants without the target condition	
	19	Report a cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, report the distribution of the test results by the results of the reference standard	
	20	Report any adverse events from performing the index test or the reference standard	
Estimates	21	Report estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals)	
	22	Report how indeterminate results, missing responses, and outliers of index tests were handled	
	23	Report estimates of variability of diagnostic accuracy between readers, centers, or subgroups of participants, if done	
	24	Report estimates of test reproducibility (e.g., interrater reliability) if done	
Discussion	25	Discuss the clinical applicability of the study findings	

Note. STARD = Standards for Reporting of Diagnostic Accuracy; MeSH = medical subject headings.

readable guide to diagnostic and screening tests prepared by David Streiner (this issue). He begins by describing the differences between tests used for screening purposes and for diagnostic purposes. Because the base rate of the target condition (e.g., an Axis I diagnosis) is often dramatically different when tests are used in these two contexts, Streiner illustrates the substantial impact that base rates have on test accuracy. Equally important, he provides a comprehensive and accessible overview of the many diagnostic efficiency statistics that can be derived from a 2 × 2 classification table, including sensitivity, specificity, positive and negative predictive power, incremental positive and negative predictive power, kappa, phi, the odds ratio, and the likelihood ratio. When describing these various statistics, Streiner also clearly indicates which measures are sensitive to the base rate of the condition being studied and which are not.

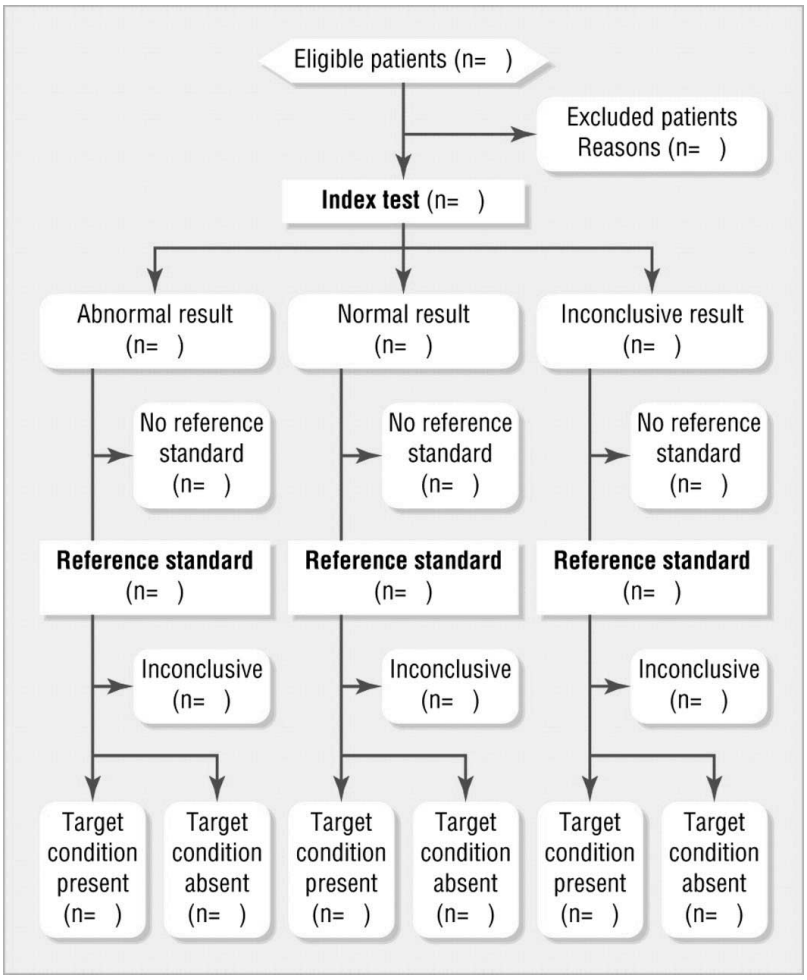

FIGURE 1 Standards for Reporting of Diagnostic Accuracy (STARD) prototype flow diagram for studies examining the diagnostic accuracy of a test.

Armed with this article and the STARD guidelines, I am confident that researchers who publish diagnostic studies on test validity in *JPA* will prepare more accurate and sophisticated contributions to the literature.

REFERENCES

Lijmer, J. C., Mol, B. W., Heisterkamp, S., Bonsel, G. J., Prins, M. H., van der Meulen, J. H. P., et al. (1999). Empirical evidence of design-related bias in studies of diagnostic tests. *Journal of the American Medical Association, 282,* 1061–1066.

Gregory J. Meyer
Department of Psychology
University of Toledo
Mail Stop 948
Toledo, OH 43606–3390
E-mail: gregory.meyer@toledo.edu

Received June 3, 2003