The STARD initiative

Derek Richards
Director, Centre for Evidence-based Dentistry, Oxford, UK

In a previous issue of the journal, we looked at a number of internationally agreed consensus guidelines on improving the quality of reporting of different study designs. These are CONSORT (Consolidated Standards of Reporting Trials), QUOROM (quality of reports of meta-analyses of randomised-controlled trials) and MOOSE (meta-analyses of observational studies). All of these guidelines are available on the CONSORT website (www.consort-statement.org). In January of this year the first official publication of STARD (Standards for Reporting of Diagnostic Accuracy) took place. This can also be found on the CONSORT website (www.consort-statement.org\stardstatement.htm). The objective of the STARD initiative is to improve the quality of reporting of studies of diagnostic accuracy.

New diagnostic tests are being developed at an increasing rate and the technology used in existing tests is continually being improved. Although the number of diagnostic tests available in dentistry is nowhere near as great as in medicine, exaggerated and biased results from poorly designed and reported diagnostic studies can trigger their premature dissemination and lead dentists into making incorrect treatment decisions. Rigorous evaluation of diagnostic tests before introduction into clinical practice could not only reduce the number of unwanted clinical consequences related to misleading estimates of test accuracy, but also limit health care costs by preventing unnecessary testing.

As part of this rigorous evaluation, studies to determine the diagnostic accuracy of a test are vital. In 1995, a paper surveying studies of diagnostic accuracy revealed that the methodological quality was at best mediocre. Assessments were hampered, however, because many reports lacked information on key elements of design, conduct and analysis of diagnostic studies,[1] a fact confirmed in other studies.[2,3] This substandard reporting of diagnostic test evaluations was discussed at the 1999 Cochrane Colloquium meeting in Rome by the Cochrane Diagnostic and Screening Test Methods Working Group. The discussion participants decided that, following the success of the CONSORT Initiative, they should develop a checklist of items that should be included in the report of a study of diagnostic accuracy. The result was the publication in January 2003 of a 25-item checklist (Table 1) based on evidence whenever it was available, together with a prototype flow diagram providing information about the method of recruitment of patients, the order of test execution and the numbers of patients undergoing the test under evaluation, the reference standard, or both the test and reference standard.

The guiding principle behind the STARD checklist was to select items that would help readers to judge the potential for bias in the study and to appraise the applicability of the findings. Two other general considerations shaped the content and format of the checklist. First, the STARD group believes that one general
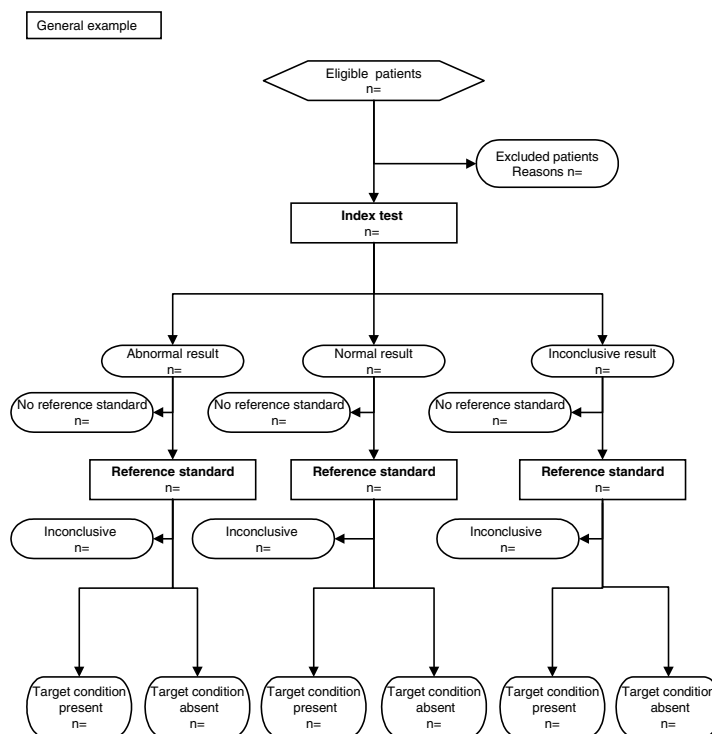
Flow Diagram

Table 1. STARD checklist for the reporting of studies of diagnostic accuracy.

Section and topic	Item number	Guideline
Title/abstract/keywords	1	Identify the article as a study of diagnostic accuracy (recommend MeSH (Medical Subject heading) heading ''sensitivity and specificity'')
Introduction	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups
Methods		
Participants	3	Describe the study population: inclusion and exclusion criteria, setting and locations where the data were collected
	4	Describe participant recruitment: was recruitment 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 the study population a consecutive series of participants defined by the 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 test 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, and/or cite references for index tests and reference standard
	9	Describe definition of and rationale for the units, cutoffs and/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	Describe whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers
Statistical methods	12	Describe methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (eg, 95% confidence intervals)
	13	Describe methods for calculating test reproducibility, if carried out
Results		
Participants	14	Report when study was carried out, including beginning and ending dates of recruitment
	15	Report clinical and demographic characteristics of the study population (eg, age, sex, spectrum of presenting symptoms, comorbidity, current treatments, recruitment centres)
	16	Report the number of participants satisfying the criteria for inclusion who did/did not undergo the index tests and/or the reference standard; describe why participants failed to receive either test (a flow diagram is strongly recommended)
Test results	17	Report time interval from the index tests to the reference standard, and any treatment administered between
	18	Report distribution of severity of disease (define criteria) in those with the target condition; 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, the distribution of the test results by the results of the reference standard
	20	Report any adverse events from performing the index tests or the reference standard
Estimates	21	Report estimates of diagnostic accuracy and measures of statistical uncertainty (eg, 95% confidence intervals)
	22	Report how indeterminate results, missing responses and outliers of the index tests were handled
	23	Report estimates of variability of diagnostic accuracy between subgroups of participants, readers or centres, if done
	24	Report estimates of test reproducibility, if done
Discussion	25	Discuss the clinical applicability of the study findings

checklist for studies of diagnostic accuracy, rather than different checklists for each specialty, is likely to be more widely disseminated and perhaps accepted by authors, peer reviewers and journal editors.

The STARD group plans to measure the impact of the statement on the quality of published reports on diagnostic accuracy using a before-and-after assessment. The group will also be providing updates when new evidence on sources of bias or variability becomes available. They also welcome comments on the current version. It will be interesting to observe whether the STARD imitative is taken up by dentistry, as it should be, or whether we will be as slow adopting this quality standard as we have been with CON-SORT.

1. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. J Am Med Assoc 1995; 274:645–651.
2. Nelemans PJ, Leiner T, de Vet HCW, van Engelshoven JMA. Peripheral arterial disease: metaanalysis of the diagnostic performance of MR angiography. Radiology 2000; 217:105–114.
3. Devries SO, Hunink MGM, Polak JF. Summary receiver operating characteristic curves as a technique for meta-analysis of the diagnostic performance of duplex ultrasonography in peripheral arterial disease. Acad Radiol 1996; 3:361–369.