The Quality of Reporting in Diagnostic Test Research: Getting Better, Still Not Optimal

In this era of evidence-based medicine, clinicians and other decision-makers turn to the scientific literature for high-quality evidence about the usefulness, precision, and accuracy of diagnostic tests. Such evidence is needed more than ever because the list of diagnostic tests is growing exponentially, and even more biomarkers, proteomics, and applications of gene expression profiling will be added in the years to come.

Studies of diagnostic accuracy can provide the necessary data. Rigorous methodologic standards in research about diagnostic tests have been slower to develop than standards for therapy studies. During the past decade, our knowledge about study design features that are associated with bias and lack of applicability in diagnostic studies has grown. Diagnostic studies with deficiencies in specific design features have been shown to be associated with biased, optimistic estimates of diagnostic accuracy compared with studies without such deficiencies (1).

Given this potential for bias, it is of paramount importance that study reports include a proper description of the study methods, in particular those design features that have been most clearly associated with bias. Within the spirit of evidence-based medicine, readers should take these features into account when examining a study and its results, pondering the decision on whether to implement changes in practice based on conclusions from the study.

Unfortunately, the quality of reporting of studies of diagnostic accuracy is often poor, making judgments about validity, bias, and applicability to patients in clinical settings difficult. The poor quality of reporting has been documented and lamented many times. One of the best known examples is a study by Reid et al. (2), who showed that even in high-impact medical journals many design features and patient characteristics were not sufficiently described.

This issue of the Journal contains an article by Lumbreras-Lacarra et al. (3) that shows the results of a study on the quality of reporting in Clinical Chemistry and Clinical Chemistry and Laboratory Medicine. For papers published in 1996, the results in terms of completeness of reporting were found to be comparable to those of the Reid study. This means that Clinical Chemistry and Clinical Chemistry and Laboratory Medicine performed as well—or as badly—as the general clinical journals in that same time window. No less than 14 of 18 papers did not describe the eligibility criteria for the study, a similar number failed to report the avoidance of review bias, and only 4 reported measures of imprecision.

Over the past decades many journals have recognized the importance of quality of reporting, and several initiatives have been developed to improve the efficiency of scientific communication. This evolution has brought us the structured paper, the “Vancouver” Uniform Requirements for Manuscripts Submitted to Biomedical Journals (4), and the structured abstract (5). A more recent phenomenon has been the development of checklists for specific study types. The first was the development of the Consolidated Standards of Reporting Trials (CONSORT) (6). This is a single-page checklist to be used by authors and reviewers to guarantee that essential features of the design and results of randomized clinical trials are well reported. The CONSORT statement has been adopted by the International Committee of Medical Journal Editors, the Council of Science Editors, and the World Association of Medical Editors.

In 1996, a group of clinical chemists developed a list of items to include in a checklist for studies of diagnostic accuracy. This list was first published in Clinical Chemistry for comment, and an amended version was published in the Journal in 2000 and incorporated in the Information for Authors (7). The checklist has been used in the review of most studies of diagnostic accuracy published in the Journal during 2001 and 2002 (personal communication, David E. Bruns, University of Virginia Medical Center, Charlottesville, VA). In the first issue of 2003, this Journal published a more general checklist, the Standards for the Reporting of Diagnostic Accuracy studies (STARD) (8, 9). This same STARD statement has been published in more than a dozen other international journals, accompanied in many of them by editorials and by the editorial decision to include STARD into the instructions for authors.

Does a checklist make a difference? Papers in journals that promoted CONSORT (BMJ, JAMA, and Lancet) showed greater improvement in quality of reporting than did papers in a journal that did not advocate its use (New England Journal of Medicine) (10). Many feel that the reporting of diagnostic studies has also improved, although some of this may be attributable to a growing sophistication among researchers in general. The study by Lumbreras-Lacarra et al. (3) provides additional evidence to suggest a role for checklists. In addition to their analysis of the papers published in 1996, these authors present an overview of papers published in Clinical Chemistry and in Clinical Chemistry and Laboratory Medicine in 2001 and 2002. Although the numbers are small, their analysis shows clear signs of improvement between 1996 and 2002 for Clinical Chemistry, with the average number of desired features increasing from two to four. A similar improvement could not yet be observed for Clinical Chemistry and Laboratory Medicine, which published the STARD statement this year, but did not use a similar checklist before (11).

Although these are encouraging results, referees and editors of journals need to do a better job in ensuring that their authors implement the STARD recommendations. For example, journals still receive—and publish—papers that fail to report on the sex and age distributions of the study participants, papers that do not report on the eligibility criteria, and papers that do a poor job in
describing how and when the diagnosis was verified. All of this matters in judging both the potential for bias and the applicability of study findings.

On a previous occasion, the editor of this Journal pointed out that a checklist works primarily by reminding authors to add information that often strengthens their conclusions but has been omitted (12). A checklist can “ensure clear and transparent reporting of studies, which will, as always, depend for their importance on the creativity and insights and effort of their authors” (12). As authors, reviewers, and readers, we can improve the quality of reporting, and hence the quality of the evidence base, for diagnostic tests. In the end, better reporting of the results of diagnostic accuracy studies has the promise to improve the efficiency and quality of healthcare.

References


Patrick M.M. Bossuyt

Academic Medical Center
University of Amsterdam
Amsterdam, NA 1105 AZ The Netherlands
Fax 31-31-20-6912683
E-mail p.m.bossuyt@amc.uva.nl

DOI: 10.1373/clinchem.2003.029736