Complete and accurate reporting of the design and implementation of studies is an essential component in guiding decisions in healthcare. Transparent reporting of studies that evaluate the accuracy of medical tests is recommended because such studies are more prone to bias than some other forms of research. Given that diagnostic accuracy is not a fixed property of tests, readers should know where and for whom the test was evaluated (1). Poor reporting of accuracy studies thus impedes the objective appraisal of such studies and limits the generalizability of study results. Suboptimal descriptions of study findings may also make tests seem more favorable than they really are, which eventually may lead to premature adoption of medical tests into practice, unnecessary testing, and high healthcare costs (2).

Ten years have passed since the initial publication of the Standards for Reporting of Diagnostic Accuracy (STARD)2 initiative. Inspired by the Consolidated Standards for the Reporting of Trials (CONSORT) Statement, the STARD statement was developed with the aim of improving the reporting of diagnostic-accuracy studies. It consists of a checklist of 25 items that ought to be reported and recommends the use of a flow diagram that describes the design of the study and the flow of patients in the study. STARD was created by a group of experts, researchers, editors, methodologists, and members of professional organizations. The STARD statement was first published in 2003 in 13 biomedical journals, including Clinical Chemistry. To date, >200 biomedical journals have encouraged the use of STARD in their instructions for authors, and the statement has been cited approximately 1654 times (ISI Web of Knowledge, February 2013). More information on the STARD initiative can be found at http://www.stard-statement.org.

After 10 years, how can we best describe the impact of STARD? Studies have evaluated its effect on the quality of reporting in various medical specialties. The general consensus is that the quality of reporting is improving, albeit slowly (3). The most recent publication by Selman and colleagues (4) evaluated the impact of STARD on the quality of reporting test-accuracy studies in obstetrics and gynecology. The authors reviewed reports published between the years 1977 and 2007 and assessed the quality of reporting before and after STARD was initiated. More than half of the STARD items were reported in 50% of the evaluated reports. Specifically, 62% of the items were reported in obstetrics, and 52% were reported in gynecology. The quality of reporting in obstetrics studies improved significantly ($P = 0.0004$) after the introduction of STARD, whereas there was no significant improvement in the reporting of gynecology studies ($P = 0.08$). Examples of the more poorly reported or missing items include: description of where and how study participants were identified and recruited into the study, blinding of the results of the index test to those interpreting the reference standard, assessment of test reproducibility, tabulation of results, and description of adverse events.

Various reasons can be postulated to explain the slow adoption of STARD. First, guidelines take a while to be adopted by journals and authors. A study by Hopewell and colleagues (5) assessed the endorsement of the CONSORT Statement in 2008 in medical journals with a high impact factor. These investigators found that after more than 10 years since the introduction of CONSORT, only 38% of high-impact journals mentioned CONSORT in their online guidelines to authors. Of these journals, a majority (63%) were not explicit in their instructions. Comparable variability and vagueness in journals’ instructions to authors for incorporating STARD have also been documented (6).

The evidence provided reveals that there is still ample room for improvement in the reporting of diagnostic-accuracy studies. With an increase in the awareness of the potential biases in diagnostic-accuracy studies, authors and reviewers of these studies will increasingly use and refer to STARD in their manuscripts.

In light of STARD’s impact thus far, what is next? Guidelines are not static, and they need to be revised or extended to accommodate a broader scope or the
changing landscape of a research field. For instance, the CONSORT Statement has been revised twice since its initial publication in 1996—in 2001 and 2010—and it has been extended >8 times (7). The STARD statement, on the other hand, has not been changed since its initial publication. Extensions to guide reporting of broader aspects of test evaluation can be considered. To start with, other elements of test evaluation in addition to diagnostic accuracy, such as analytical validity, also need to be reported clearly. Analytical validity aims at assessing the accuracy of a test for measuring the concentration or changes in the concentration of a substance (8), and such an evaluation usually marks the initial stage of assessing tests such as laboratory tests and biomarker measurements. Reporting items in the current STARD statement, such as those related to the reference standard, adverse events, and clinical applicability of tests may not be applicable. In such cases, these items will be excluded in the STARD extension for tests used for monitoring changes in circulating concentrations.

Second, STARD may need to be extended to allow its application to predictive or prognostic tests or biomarker measurements. The development of markers and tests to help predict a future state or event (e.g., recurrence, progression, or response to therapy) has recently gained momentum. Reporting recommendations (REMARK) have been developed to guide the reporting of prognostic studies of tumor markers (9). Criteria for the evaluation of novel markers of cardiovascular risk have been published, but they have not yet been accompanied by guidelines for transparent and complete reporting.

Similar to the CONSORT extension for abstracts, another priority for STARD is extending it to provide guidelines for the reporting of abstracts of test-accuracy studies (10). Abstracts provide an overview of a study, and decisions may be made on abstracts alone when full texts are not available. Hence, abstracts need to be reported well so that they accurately reflect the results of the main study.

Protocols serve as a foundation for the proper execution of a study and the reporting of its results. To limit variation and incomplete reporting, guidelines are required to standardize the reporting of items in a protocol. For interventions, the SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) initiative has been published (11). Unlike clinical trials, however, formal registration of protocols of diagnostic-accuracy or biomarker-evaluation studies is not yet an official requirement. Consequently, that fact dampens the impetus for authors of test-accuracy studies to be keen on what to include in the protocols. Calls have been made to open a register for test and biomarker evaluation (12). If this register comes to fruition, STARD may be extended to guide protocol reporting for tests and biomarkers.

The STARD steering group also plans to revise the style of language used in the STARD statement so that it is consistent with CONSORT and other, more recent, guidelines.

Other potential plans for STARD include making the guidelines more understandable to researchers involved in test-accuracy evaluations. One way would entail revamping the STARD website to include more information or presentation graphics. Such graphics could include PowerPoint presentations of talks related to STARD or test accuracy and more examples of good and poor flow diagrams of accuracy studies. Additionally, templates to guide authors in the use of STARD and discussion forums or blogs where researchers can ask questions or share experiences can be incorporated into the website.

Ten years on, STARD has slowly improved the quality of the reporting of diagnostic-accuracy studies. We anticipate that the quality of reporting will continue to improve as more readers become aware of STARD and as more journals become explicit in their instructions to authors regarding its use. The quality of reporting should also improve if STARD is modified so that it is accompanied by extensions and comparable guidelines for the evaluation of other features of medical tests, and if the STARD website is revamped to make its content more comprehensible to users.

References

5. Hopewell S, Altman DG, Moher D, Schulz KF. Endorsement of the CONSORT Statement by high impact factor medical journals: a survey of jour-


