Special Report

Call for Action: Journals Need to Insist on Full Reporting of the Analytical Characteristics of Biomarkers

Zhen Zhao, PhD,1 David B. Sacks, MB, ChB2*

Laboratory Medicine 2021;52:7-9
DOI: 10.1093/labmed/lmaa097

Keywords: biomarker, analytical characteristics, guideline, reporting, reproducibility, transparency

There is a growing movement to encourage the publication of complete methods and to report the quality of data in basic research.1-3 High-profile journals such as Science and Nature have published and endorsed several guidelines and checklists to promote transparency and reproducibility.3-5 Although these guidelines place considerable emphasis on the transparent reporting of statistics, analytical validation has not been addressed. It is important to be aware that failure to adhere to appropriate methodological standards can lead to biased and misleading results, even when suitable statistical methods are applied.6,7

The lack of transparent, adequate reporting of analytical method validation may undermine credibility and reduce the relevance of a clinical study. Moreover, analytical shortcomings may be a barrier to generating reproducible results and translating important initial findings into clinical practice. Nevertheless, there has been no objective information available on the extent of characterization or reporting of the analytical performance of chemical biomarkers in published clinical studies. Therefore, we assessed the reporting of analytical characteristics of biomarkers used in clinical research and evaluated the extent of reported data on assay precision.8

We analyzed all the clinical studies that were published in 5 high-impact medical journals: the Annals of Internal Medicine, JAMA: The Journal of the American Medical Association, The Lancet, The New England Journal of Medicine, and PLOS Medicine, over a 10-year period. Using the term “biomarker” as a search criterion, we identified 544 studies (with 1299 biomarkers) in which biomarkers were used for patient classification or inclusion/exclusion criteria or as a study outcome. For each study, we systematically reviewed the reporting of 11 analytical validation characteristics and the quality of validation. The key characteristics included analytical accuracy or trueness, imprecision, analytical sensitivity, interferences, reportable range, reference interval, cutoff or decision limits, quality control, and calibration.

Remarkably, our analysis revealed that the median reporting rate was zero. No characteristic was provided for 67% of the biomarker measurements, and more than 4 characteristics were reported for only 3% of the biomarkers. We also examined the completeness, comprehensiveness and quality of the reporting using total imprecision as an example because it showed the highest reporting rate at 13%. Of the 274 articles that reported imprecision, 76% reported only a total coefficient of variation, with no details about the concentrations measured or the number of days of testing. These findings suggest that even for the little information that was reported, the quality of method validation was far from adequate to meet expectations for the clinical use of biomarkers.

Values for a given patient sample can vary significantly between assays from different manufacturers (even for methods approved by the U.S. Food & Drug Administration) because of a lack of assay harmonization. Nevertheless, a persistent

Abbreviations:
FDA, U.S. Food & Drug Administration; STARD, Standards for Reporting of Diagnostic Accuracy.

1Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, New York, 2Department of Laboratory Medicine, National Institutes of Health, Bethesda, Maryland

*To whom correspondence should be addressed.
david.sacks2@nih.gov
problem we observed was a failure to report the name of the manufacturer, which was detected for 53% of biomarker measurements. As an illustration, we carefully examined the 2 biomarkers most commonly measured in the publications we assessed. The manufacturers of cardiac troponin (63 reports; 5%) and glucose (47 reports; 4%) methods were stated in only 35% and 33% of the studies, respectively. Cardiac troponin I results can vary up to 33-fold among assays from different manufacturers. Similarly, glucose results can vary substantially among methods, especially between point-of-care glucose meters and central laboratory methods. Moreover, even if the manufacturer remains unchanged, methods can change over time with improved and/or more standardized measurements. Clinical studies can last many years, be performed at multiple locations, and use methods from different manufacturers. Therefore, documenting the assay method and keeping track of any changes to the method are essential to ensure the consistency of the measurement over the course of a study.

Analytical method validation by clinical laboratories is required in the United States by the Clinical Laboratory Improvement Amendments. Unfortunately, this concept has received minimal attention in the general scientific community. In basic and clinical research settings, investigators often assume that analytical validation and method optimization are the manufacturers’ responsibility. However, this guideline pertains only to methods that have been approved by the FDA. Only 24% of the biomarker measurements in our study used FDA-approved methods. “Research Use Only” assays, which have not been cleared or approved by the FDA, were commonly used in the publications we examined. Although they are acceptable for clinical research, it is important to realize that these assays have not undergone analytical validation as rigorous as those of FDA-approved methods. Therefore, it is essential that the pertinent validation be performed before the assays are used in clinical studies.

In 2003, the Standards for Reporting of Diagnostic Accuracy (STARD) Initiative developed a checklist for the complete and accurate reporting of diagnostic accuracy studies. However, emphasis in STARD is placed on the reporting of clinical rather than analytical performance of methods. Clinical performance pertains to how results correctly classify patients and correlate with a patient’s disease state and outcome. By contrast, analytical validation ensures an accurate and reproducible measurement, which serves as the foundation of clinical evaluation. The only analytical validation item in STARD is cutoff. Notwithstanding this requirement, our evaluation revealed that only 6% of the 1299 biomarker measurements we assessed reported the cutoff. This poor adherence indicates that journals have not implemented STARD in the editorial and peer-review process as rigorously as they should. Our findings are congruent with 2 recent reports suggesting that although more than 200 biomedical journals have endorsed STARD, adherence to guidelines remains suboptimal.

Although multiple factors contribute to the lack of reproducibility of clinical research, our findings specifically highlight how inadequate reporting of analytical characteristics can create barriers to study replication. An important contribution of our work is the opportunity to challenge the field to move beyond the current requirements of reporting the analytical items used in clinical studies. It has been suggested that standards for design and measurement would make clinical trials reproducible and usable. Because numerous biomarker measurements have not been standardized, we recommend that when biomarkers are defined in clinical guidelines and used in clinical trials, basic analytical information such as cutoff, methods, and manufacturers should be provided to facilitate the interpretation and generalization of results to clinical practice. Although word limits could preclude detailed information regarding the methods, this information could readily be included in supplemental materials.

We also encourage enhanced communication between clinical laboratory staff and clinicians to raise awareness of the underreporting issues. The clinical community should come together to provide consensus recommendations on more informative reporting standards for analytical methods and biomarkers. Such guidelines could serve as a template for investigators to reduce analytical error when designing studies, for clinicians to better evaluate the analytical performance of assays before implementing them in patient care, for funding bodies to ensure the quality of data, and for other scientists to reproduce the biomarker measurements.

We recognize that there are other important items beyond the establishment and enforcement of guidelines/checklists that need to be addressed. First, research scientists often receive little training in the analytical and regulatory aspects of method development and validation. Second, there is no universal approach to validating analytical methods. The extent of analytical rigor need not be applied equally to discovery, initial exploratory studies, clinical trials, and clinical practice. Therefore, in an effort to develop a guideline for authors, a checklist that meets the purpose of the tests should be considered, along with the FDA-approval status and quantitative nature of the tests.
In summary, we have identified a previously underappreciated deficiency in the published clinical literature. The lack of reporting of analytical parameters of biomarkers is a threat to the interpretation and replication of clinical studies. Our analysis raises awareness and highlights an important limitation of many clinical studies. “How one measures matters,” wrote Patrick Bossuyt, professor of clinical epidemiology at the University of Amsterdam, in an editorial in *Clinical Chemistry* that commented on our study. He also perceived the results of our study as a call for action for clinical laboratory staff to request sufficient details about methods when reviewing manuscripts and commenting on study protocols. Our results emphasize the need for the development of “fit-to-purpose” guidelines that are easy for researchers, funding agencies, journals, and other stakeholders to follow. Additional solutions include mentoring young investigators on analytical rigor and transparency. Implementation of these strategies should enhance thorough and transparent reporting, which will enable readers to assess the quality of the work, evaluate the validity of the results, repeat the experiments, and ultimately decrease irreproducibility in future clinical research. LM

**Acknowledgments**

DBS is funded by the Intramural Research Program of the National Institutes of Health.

**References**

1. Iqbal SA, Wallach JD, Khoury MJ, Schully SD, Icencis JP. Reproducible research practices and transparency across the biomedical literature. *PLoS Biol.* 2016;14(1):e1002333.

2. Baggerly K. Disclose all data in publications. *Nature.* 2010; 467(7314):401.

3. Nosek BA, Alter G, Banks GC, et al. Scientific standards. Promoting an open research culture. *Science.* 2015;348(6242):1422–1425.

4. Announcement: towards greater reproducibility for life-sciences research in *Nature* [editorial]. *Nature.* 2017;546(7656):8.

5. Announcement: transparency upgrade for nature journals [editorial]. *Nature.* 2017;543(7645):286.

6. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA.* 1999;282(11):1061–1066.

7. Riley RD, Abrams KR, Sutton AJ, et al. Reporting of prognostic markers: current problems and development of guidelines for evidence-based practice in the future. *Br J Cancer.* 2003;88(8):1191–1198.

8. Sun Q, Welsh KJ, Bruns DE, Sacks DB, Zhao Z. Inadequate reporting of analytical characteristics of biomarkers used in clinical research: a threat to interpretation and replication of study findings. *Clin Chem.* 2019;65(12):1554–1562.

9. Apple FS, Ler R, Murakami MM. Determination of 19 cardiac troponin I and T assay 99th percentile values from a common presumably healthy population. *Clin Chem.* 2012;58(11):1574–1581.

10. Jennings L, Van Deerlin VM, Gulley ML; College of American Pathologists Molecular Pathology Resource Committee. Recommended principles and practices for validating clinical molecular pathology tests. *Arch Pathol Lab Med.* 2009;133(5):743–755.

11. Bossuyt PM, Reitsma JB, Bruns DE, et al.; Standards for Reporting of Diagnostic Accuracy Group. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. The Standards for Reporting of Diagnostic Accuracy Group. *Croat Med J.* 2003;44(5):635–638.

12. Zhao Z, Sacks DB. Assay development in clinical pathology. In: McManus LM, Mitchell RN, eds. *Pathobiology of Human Disease: A Dynamic Encyclopedia of Disease Mechanisms.* London, UK: Elsevier; 2014:3194–3206.

13. Korevaar DA, Wang J, van Enst WA, et al. Reporting diagnostic accuracy studies: some improvements after 10 years of STARD. *Radiology.* 2015;274(3):781–789.

14. Dilauro M, McInnes MD, Korevaar DA, et al. Is there an association between STARD statement adherence and citation rate? *Radiology.* 2016;280(1):62–67.

15. Dickersin K, Mayo-Wilson E. Standards for design and measurement would make clinical research reproducible and usable. *Proc Natl Acad Sci U S A.* 2018;115(11):2590–2594.

16. Bossuyt PM. Laboratory measurement’s contribution to the replication and application crisis in clinical research. *Clin Chem.* 2019;65(12):1479–1480.