



Adherence to the Standards for Reporting of Diagnostic Accuracy Studies (STARD): a survey of four journals in laboratory medicine

Fei-Fei Zheng¹, Wei-Hong Shen¹, Fang Gong¹, Zhi-De Hu^{2,3^}, Giuseppe Lippi^{4^}, Ana-Maria Šimundić^{5^}, Patrick M. M. Bossuyt^{6^}, Mario Plebani^{7^}, Kaiping Zhang^{8,9}

¹Department of Laboratory Medicine, the Affiliated Hospital of Jiangnan University, Wuxi, China; ²Department of Laboratory Medicine, the Affiliated Hospital of Inner Mongolia Medical University, Hohhot, China; ³Journal of Laboratory and Precision Medicine Editorial Office, Guangzhou, China; ⁴Section of Clinical Biochemistry, University of Verona, Verona, Italy; ⁵Department of Medical Laboratory Diagnostics, University Hospital “Sveti Duh”, Faculty of Pharmacy and Biochemistry, Zagreb University, Zagreb, Croatia; ⁶Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands; ⁷Department of Medicine, University of Padova, and Department of Integrated Diagnostics, University-Hospital of Padova, Padova, Italy; ⁸Editorial Office, AME Publishing Company, Hong Kong, China; ⁹School of Public Health, Imperial College London, London, UK

Contributions: (I) Conception and design: FF Zheng, ZD Hu; (II) Administrative support: WH Shen, F Gong; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: FF Zheng, ZD Hu; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Zhi-De Hu. Department of Laboratory Medicine, The Affiliated Hospital of Inner Mongolia Medical University, Hohhot, China; Journal of Laboratory and Precision Medicine Editorial Office, Guangzhou, China. Email: hzdlj81@163.com.

Background: The Standards for Reporting of Diagnostic Accuracy Studies (STARD) statement has been updated in 2015. Many diagnostic test accuracy (DTA) studies have been published in medical laboratory journals, but their adherence to the updated STARD statement remains unknown.

Methods: We searched the PubMed database to verify studies published in 4 laboratory journals, including *Clinical Chemistry*, *Clinical Chemistry and Laboratory Medicine*, *Clinica Chimica Acta*, and *Clinical Biochemistry*, in 2019. DTA studies were identified and their adherence to the STARD statement was assessed.

Results: A total of 45 studies were included in this analysis. Overall, 18 out of 34 STARD items were reported. The items (adherence rate) of sample size estimation (4%), adverse events (9%), protocol (9%), registration (16%), missing value (22%), indeterminate results (18%), and cross-tabulation (22%) were the most frequently unreported items.

Conclusions: Adherence to the STARD statement in DTA articles published in laboratory medicine seems as yet unsatisfactory. Our study emphasizes the necessity to improve the reporting quality of DTA studies published in medical laboratory journals.

Keywords: STARD statement; laboratory medicine; diagnostic test accuracy (DTA)

Submitted Apr 05, 2021. Accepted for publication May 23, 2021.

doi: 10.21037/atm-21-1665

View this article at: <http://dx.doi.org/10.21037/atm-21-1665>

[^] ORCID: Zhi-De Hu, 0000-0003-3679-4992; Giuseppe Lippi, 0000-0001-9523-9054; Ana-Maria Šimundić, 0000-0002-2391-5241; Patrick M. M. Bossuyt, 0000-0003-4427-0128; Mario Plebani, 0000-0002-0270-1711.

Introduction

Diagnosis is a crucial step in disease management. A diagnostic test accuracy (DTA) study is a common type of research design that aims to assess the diagnostic accuracy of index tests. Unlike interventional studies, in which efficiency and safety of interventional approaches and controls are compared, DTA studies assess the diagnostic accuracy of an index test against the reference standard. Sensitivity, specificity, positive and negative likelihood ratio (PLR/NLR), positive and negative predictive value (PPV and NPV), and diagnostic odds ratio (DOR) are basic metrics for DTA (1).

Biomarker assessment plays an important role in diagnosing most human diseases (2). The diagnostic accuracy of a biomarker needs to be adequately assessed before it can be introduced into clinical practice. Complete and accurate reporting of DTA studies is also critical for reporting the applicability of findings and conclusions.

It is well-known that a systematic review or meta-analysis of DTA studies represents a high evidence level and is recommended by guidelines. However, incomplete and inaccurate reporting usually makes the systematic reviewer unable to assess the quality of the studies and of the data included in pooled analyses or meta-analyses. With the specific aim of improving the quality of reporting in DTA studies, the Standards for Reporting of Diagnostic Accuracy Studies (STARD) statement was released in 2003 (3,4). The 2003 version of the STARD statement was updated in 2015, with the number of reporting items being increased from 25 to 34 (5,6).

An earlier report investigated DTA studies published in medical laboratory journals and their adherence to the 2003 version of STARD (7). To the best of our knowledge, there is no published evaluation concerning the adherence to the updated STARD statement in medical laboratory journals. We conducted this study with the aim of surveying to what extent recent articles published in medical laboratory journals adhere to the STARD 2015 statement.

Methods

Literature retrieval

We search the PubMed database to identify potentially eligible studies published in 4 laboratory medicine journals indexed in the database of the United States National Library of Medicine at the National Institutes of Health: *Clinical Chemistry*, *Clinical Chemistry and Laboratory*

Medicine, *Clinica Chimica Acta*, and *Clinical Biochemistry*. The search algorithm was as follows: (("2019/01/01"[ppdat]: "2019/12/31"[ppdat]) and (clin chem[journal] or clin chem lab med[journal] or clin chim acta[journal] or clin biochem[journal]) and (ROC Curve[mesh] or Sensitivity and Specificity[mesh] or sensitivity[Title/Abstract] or specificity[Title/Abstract] or diagnos*[ti])) not (editorial[pt] or review[pt] or Systematic Review[pt] or comment[pt]).

Inclusion and exclusion criteria

All articles reporting the diagnostic accuracy of biomarkers were eligible for our analysis. The exclusion criteria were as follows: (I) diagnostic accuracy of an index test was reported the focus of the study was on technical development rather than diagnostic accuracy; (II) original studies published as letters; (III) comments, editorials, conference abstracts, or reviews.

Two reviewers independently screened the identified articles to verify eligibility. In the first round, the title and abstract were screened to exclude the irrelevant items. In the second round, the full text of the article was read, and eligibility was further verified. Any disagreements in selection were resolved by consensus.

Reporting quality assessment and data extraction

Two reviewers independently assessed reporting of included articles according to the STARD guideline. In addition, information concerning the country, type of data collection (prospective or retrospective), field (e.g., cancer, cardiovascular disease, autoimmune disease), center (single center or multicenter) was also extracted. Any disagreements in data extraction or in completeness of reporting were resolved by consensus. We calculated the percentage of included studies reporting each of the STARD items.

Statistical analysis

All analyses were performed in excel version 2019 (Microsoft). Categorical data were expressed as absolute and relative frequencies.

Results

Summary of eligible studies

Figure 1 shows a flowchart of the selection process.

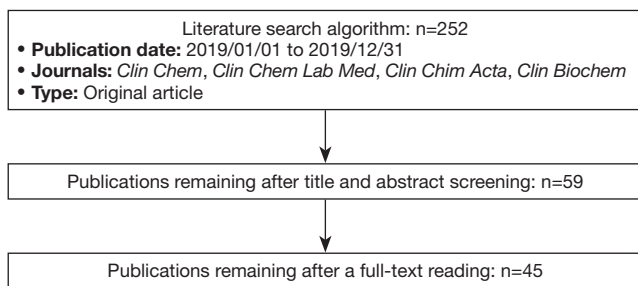


Figure 1 Flowchart of study selection.

Ultimately, 45 articles were included. The characteristics of the corresponding studies are summarized in *Table 1*.

Adherence to the STARD guideline

Figure 2 summarizes adherence to STARD guidelines for the included articles. Only 1 of the 45 articles (4%) reported the method of sample size estimation and 2 studies (9%) reported adverse events of the index test. Furthermore, full protocol availability (9%), registration number (16%), indeterminate index test results (18%), missing data (22%), and cross-tabulation (22%) were the least frequently reported items.

Furthermore, 18 (40.0%) of the articles included 15 or fewer STARD items, and 17 (37.8%) reported 16–23 items. The remaining 10 (22.2%) articles reported over 24 items. The average number of reported items was ~18 (range, 7–33).

Discussion

To the best of our knowledge, this is the first study investigating the adherence to the 2015 STARD of DTA studies published in medical laboratory journals. In keeping with the results of previous investigations on point-of-care ultrasound research (8) and magnetic resonance imaging (9), we found that the overall reporting quality of DTA articles published in medical laboratory journals was unsatisfactory. For example, the lowest number of reported items was 7, which means that the average study report includes only 20% of the STARD items. The most common unreported item was sample size estimation (reported in only 1 study), followed by adverse events, and protocol.

As many studies reported in medical laboratory journals focus on the diagnostic accuracy of biomarkers, those included in this work may be considered representative of biomarker-related DTA studies. Nonetheless, the

Table 1 Summary of eligible studies

Characteristic	Number (%)
Country	
Total	45 (100)
China	10 (22.2)
Italy	6 (13.3)
Brazil	3 (6.7)
Spain	3 (6.7)
Germany	3 (6.7)
USA	2 (4.4)
Japan	2 (4.4)
Others	16 (35.6)
Target diseases	
Cancer	11 (24.4)
Cardiovascular	9 (20.0)
Infectious disease or sepsis	6 (13.3)
Autoimmune disease	5 (11.1)
Others	14 (31.1)
The STARD checklist submitted	
Yes	6 (13.3)
No	39 (86.7)
Data collection	
Prospective	13 (28.9)
Retrospective	16 (35.6)
Not reported	16 (35.6)
Center	
Single center	32 (71.1)
Multicenter	13 (28.9)
Journal	
<i>Clin Chem</i>	12 (26.7)
<i>Clin Chem Lab Med</i>	8 (17.8)
<i>Clin Chim Acta</i>	16 (35.6)
<i>Clin Biochem</i>	9 (20.0)

conclusions of our analysis may not be straightforwardly translated to DTA studies investigating imaging, instrumental, or pathological tests, as the manner in which these are conducted is inherently different and involved the

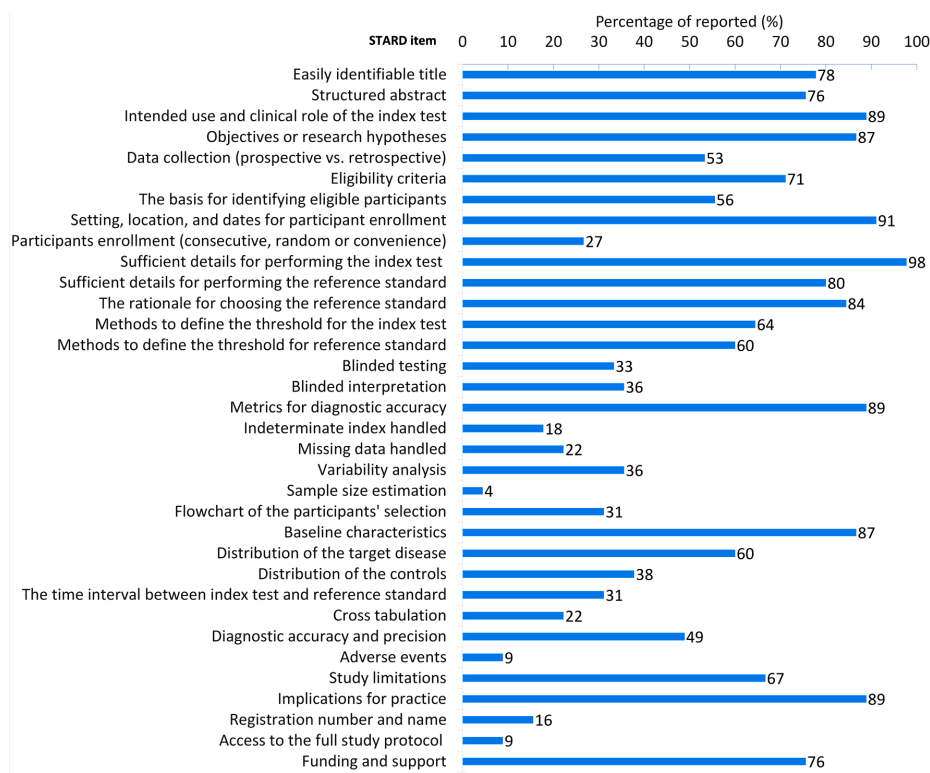


Figure 2 Adherence to STARD statement of the DTA studies published in laboratory medicine journals.

patients themselves and not simply their biological samples.

Complete and accurate reporting is crucial when performing systematic literature reviews and meta-analyses because the reported items relate to either version [2003 (10) or 2011 (11)] of the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool. Target population, index test, and reference standard are three basic elements for DTA. If these elements are not clearly reported, the quality of the DTA study cannot be assessed. As concerns the target population, a study failing to report whether the participants were enrolled consecutively is labelled as *unknown* in the patient selection domain in QUADAS-2. This means that whether or not there is a bias in patient selection or to what extent the results may be accurate and reproducible cannot be established. Notably, we found that only one-fourth of the DTA studies reported whether the participants were consecutively enrolled. In addition, only a small portion of DTA studies reported how missing and indeterminate data would be handled, which should be an essential element of DTA studies. These two items are crucial for assessing to what extent the study sample is representative.

The majority of index tests in articles published in laboratory medicine focus on biomarkers, and measurements

data are typically expressed as continuous variables. For continuous variables, a trade-off exists between sensitivity and specificity. Therefore, a clear and reasonable definition of the threshold is essential for calculating the diagnostic sensitivity and specificity. Only two-thirds of eligible studies reported a clear threshold value. Moreover, some articles did not indicate clearly in the statistical section whether a prespecified or data-driven threshold was adopted. This means that the index test domain in the QUADAS-2 tool should be labeled as unknown in quality assessment.

Some studies did not report whether all samples were tested with the same reference technique, making it impossible to establish whether partial verification or different verification bias could have adversely influenced the study. Some DTA studies did not even report the time of specimen collection, making it impossible to ascertain the stability of the analyte and if different intervals were used for conducting the index and the reference tests.

Taken together, incomplete and inaccurate reporting will hinder the generalization of DTA studies. Generally, the STARD statement was not adequately followed in DTA studies published in laboratory medicine journals, and thus actions should be taken to improve the reporting quality of

DTA published studies.

Questions to be further considered

What are the possible reasons for the low rate of reporting sample size estimation?

Expert opinion: Ana-Maria Šimundić

It is difficult to say. Probably a lack of knowledge and understanding of the importance of power analysis and how this should be calculated.

Expert opinion: Karel G. M. Moons

This is because, as compared to randomized therapeutic studies, there are no good or widely accepted guidelines for estimating required sample size for obtaining a desired sensitivity, and specificity of predictive values with corresponding confidence interval (CI). Moreover, books or guidance papers for DTA studies do not stress the need for a priori sample size estimations. You might have a look at the paper of Simel (12).

Expert opinion: Mario Plebani

Many reasons: lack of awareness of the STARD existence and poor compliance with the STARD checklist. In addition, most studies on IVD tests and biomarkers do not receive significant support, such as that provided for drug evaluation. This, in turn, makes it difficult to collect an appropriate number of samples/patients to be evaluated, and therefore sample size estimation follows a more “practical” than scientific approach.

Expert opinion: Patrick M. M. Bossuyt

This is due to a combination of factors. A lack of awareness from the study authors may be one of the main reasons, and peer reviewers, funders and regulators have not asked for sample size justifications, though simple methods exist. See the reference (13).

Expert opinion: Giuseppe Lippi

There are many reasons for this:

- (I) Ignorance of the existence of STARD.
- (II) Disagreement with one or more items on the STARD checklist.
- (III) Unfeasibility in applying all STARD items to the biomarker in evaluation (e.g., STARD cannot be applied as a whole to the evaluation of molecular biomarkers).

- (IV) Incompleteness of experiments (due to shortage of time/reagents/money) to fulfill all STARD items.
- (V) Not enough space to report all items, when the manuscript is submitted as a “letter to the editor” or a “short communication”.

Editor opinion: Kaiping Zhang

There are many reasons for this poor reporting, including the study’s design and conduct before it is completed and the study’s reporting after it is completed. Either way, the journal editors and the reviewers are essential “gatekeepers”. Some authors may not be aware of the importance and necessity of reporting these items. One of the reasons may be that the journal’s requirements are not strict enough. Although three of the four journals mention the STARD guideline in their Author Guidance, we can see from the data in *Table 1* that only six articles (13.3%) submitted the STARD checklist. Thus, it is not easy to know whether the editorial office has required or strictly checked these items one by one. The reporting quality would undoubtedly be significantly improved if the journals have required a STARD checklist submission, have arranged an editor to check the items one by one, and have required the author to explain in the article why some items could not be reported.

For studies investigating the diagnostic accuracy of biomarkers, blind testing seems meaningless because the level of biomarkers is determined by an instrument in an objective manner. Is it essential to report this item?

Expert opinion: Ana-Maria Šimundić

No. Blinding is only important for methods which are subjective like manual microscopy. For automated methods, blinding is of no significance.

Expert opinion: Karel G. M. Moons

Yes indeed, although many biomarkers in daily practice are subsequently interpreted and sometimes even dichotomized for decision making. Then it is handy and an extra comfort that blinding was achieved. Note that there are two type of blinding:

Blinding the observer of the outcome (reference test) for the results of the index test (biomarker), which has to do with the objectivity of the ref test result.

Blinding the observer of the index test for the results of the ref test, which has to do with the objectivity of the index test result and which is automatically fulfilled with a prospective cohort design. Please see our paper (14).

Expert opinion: Mario Plebani

Blinding testing is totally meaningless if it refers to automated testing. What should be relevant is to assure that laboratories are working on anonymized samples and following the same procedures adopted in clinical practice for all patient samples.

Expert opinion: Patrick M. M. Bossuyt

That would depend on the context. In several cases, the measurement is not just produced by the instrument—there are preanalytical issues, measurement issues, and post-analytical issues. I would agree that the relevance of blinding, in general, depends on the level of judgment and decision-making and human interaction involved.

Expert opinion: Giuseppe Lippi

I fully agree with this statement. Blind testing is 100% meaningless. Knowing in advance what you are testing will have no impact on the outcome, as some scientists may cheat even afterwards while making the statistical analysis and/or reporting data.

Editor opinion: Kaiping Zhang

Good point. I cannot agree more with this statement. However, the item 13a and 13b in STARD aim to highlight the importance of standardized and objective testing. As a general guideline, the STARD has covered scenarios of both objective and subjective assessing. As with any other general reporting guidelines, it must be balanced in terms of representativeness, context and simplicity. In a nutshell, it is a triple-edged sword.

The majority of DTA studies did not publish their protocol in advance, and no studies with the same cohort have been published. Under such conditions, how can the items of the protocol be reported? (item 29)

Expert opinion: Ana-Maria Šimundić

If a study protocol has not been previously published, then it should be provided in the manuscript.

Expert opinion: Karel G. M. Moons

You cannot—similar to question 1—as there is hardly any emphasis on prepublishing protocols of diagnostic test or prognostic test studies. This is much more stressed and pursued now by journals like *BMC Diagnostic & Prognostic Research*. See the paper by Peat *et al.* (written for prognostic test studies but indicated that it was also applicable to

diagnostic test studies) (15).

Expert opinion: Mario Plebani

This is an intriguing issue but, as a matter of fact, many studies do not have a protocol. This, in any case, should be mentioned in the manuscript as a limitation of the study itself.

Expert opinion: Patrick M. M. Bossuyt

We are aware that, unfortunately, many studies do not have a protocol. The absence of a protocol (item 29) could be mentioned, but our hope is that item 29 will encourage the development and the release of a protocol.

Expert opinion: Giuseppe Lippi

I do not think that this will be necessary. In a fast-moving environment like lab medicine (especially in the time of COVID-19), publishing study protocols is a highly questionable waste of time.

Editor opinion: Kaiping Zhang

Publication of protocol is encouraged when circumstances permit but not mandatory. There is a massive difference between ‘publication’ and ‘can be accessed’ (the original phrasing of item 29 in STARD). The focus is to improve the transparency of the process and the metadata through available ways, e.g., provide the protocol under the data sharing statement’s approval. In summary, the authors should at least complete a protocol, no matter they plan to publish it or not. The authors and the journal editors will be jointly responsible for access to the protocol.

When you are reviewing a DTA article relevant to a biomarker, which STARD items are you most attentive to?

Expert opinion: Ana-Maria Šimundić

Item 5—study design: this is of vital importance to the quality of the results. Most problematic is the case-control study design—this causes huge bias and overestimates measures of diagnostic accuracy.

Item 9—recruitment of participant: patients have to be recruited in a subsequent series of patients. Everything else will implement bias.

Item 10b—reference standard: this is also very important. Many variations exist and they all cause bias: for example, partial verification bias (not all subjects will have the reference standard), or differential verification bias (there are two reference standards and some patients will have one

and others will have the other), etc.

Item 23—cross tabulation: this is very often missed. If provided, it gives clarity and makes results easy to understand. Also, it makes it easy to check results by recalculation.

Item 24—diagnostic accuracy and precision: the 95% CI is frequently not provided. Without the CI it is impossible to interpret measures of diagnostic accuracy.

Expert opinion: Karel G. M. Moons

I use QUADAS or PROBAST (see www.probast.org), rather than STARD.

Expert opinion: Mario Plebani

Data collection (prospective versus retrospective), reference standard and related issues, diagnostic accuracy, and limitations.

Expert opinion: Patrick M. M. Bossuyt

That depends on my role. As a reviewer, I notice that a description of the methods for recruiting study participants or study samples is very often missing, which is unjustifiable.

Expert opinion: Giuseppe Lippi

Limit of detection (LoD), limit of blank (LoB), and sample size used for comparing data with the reference method. Correct use of parametric/non-parametric testing for the statistics is also mandatory.

Editor opinion: Kaiping Zhang

Item 5 (prospective or retrospective data collection), item 6 (eligibility criteria), item 16 (missing data), item 18 (sample size), item 25 (adverse events), item 26 (study limitations) and item 27 (implications for practice).

In some DTA studies, a participant's diagnosis is complex. For example, in a study investigating the diagnostic accuracy of pleural markers for parapneumonic pleural effusion (pneumonia-related pleural effusion), patients with heart failure-induced pleural effusion are categorized as a control. If a patient has heart failure caused by pneumonia, how should the authors categorize the patient and report results?

Expert opinion: Ana-Maria Šimundić

Diagnostic criteria should be defined in advance. All subjects are categorized according to these criteria into those with (A) and those without (B) the disease. In both groups, the index

test is measured and measures are calculated. It is simple. There are no controls.

Expert opinion: Karel G. M. Moons

Misclassification of outcomes is indeed complex and should be prevented as much as possible. If one developed the target outcome (regardless of also developing any other outcome), it should be classified as the target outcome being present. Moreover, I never think of cases and controls. A case-control design is often used in lab research, but is very wrong and should be avoided. See the relevant studies (16-18).

Expert opinion: Mario Plebani

The adoption of clinical guidelines and recommended diagnostic criteria is mandatory. The reporting of the classification and rationale that guided the identification of control group(s) is very important. It should be highlighted that it is not enough to define who are included in the disease group (A) and in non-disease (B), but also possible confounding pathophysiological issues that should be present in the “real world”.

Expert opinion: Patrick M. M. Bossuyt

The aim of the STARD reporting guideline is to improve complete and transparent reporting. It should be clear to the reader how the classification was arrived at, and why that classification was the right one for the study objectives.

Expert opinion: Giuseppe Lippi

These patients should not be categorized! Patients with disease caused by other pathologies should be treated as a third group or else excluded from either group, with this being stated in the exclusion criteria.

Editor opinion: Kaiping Zhang

Item 5 (prospective or retrospective data collection), item 6 (eligibility criteria), item 16 (missing data), item 18 (sample size), item 25 (adverse events), item 26 (study limitations) and item 27 (implications for practice).

For biomarker-related DTA studies with a small sample size and a retrospective design, variability analyses seem impossible. In cases like this, how should the corresponding items be reported? (Item 17)

Expert opinion: Ana-Maria Šimundić

I suggest the following: due to the retrospective design, the

analysis of variability should not be done.

Expert opinion: Karel G. M. Moons

I am not familiar with variability analysis, but the term suggests that it is more a technical analysis than a clinical study.

Expert opinion: Mario Plebani

In this case, as an editor I would reject the paper. In any case, the inclusion of a statement such as “due to the retrospective design of the study, the analysis of variability was not performed”, should be done.

Expert opinion: Patrick M. M. Bossuyt

Unfortunately, many biomarker studies are very small in sample size. This is regrettable, as the performance of biomarkers and medical test is known to vary across subgroups. Item 17 encourages authors to preplan and report such analyses in a sufficiently powered study.

Expert opinion: Giuseppe Lippi

This should not be reported.

Editor opinion: Kaiping Zhang

Hard to report. List this as one of the limitations of the study.

In many DTA studies with a retrospective design, the index test results are not blinded to the clinician who makes the final diagnosis. In these cases, is it necessary to report the items of blindness (Item 13b)?

Expert opinion: Ana-Maria Šimundić

Only if diagnostic criteria are not well defined, and if a diagnosis cannot be made in an objective and reproducible manner.

Expert opinion: Karel G. M. Moons

No, see the paper by Moons and Grobbee (14).

Expert opinion: Mario Plebani

Yes, I strongly believe that this is mandatory, but not in the case in which the clinical diagnosis needs to be achieved using the laboratory information.

Expert opinion: Patrick M. M. Bossuyt

It should be made clear (item 13b) that the clinician making the call had access to the biomarker results.

Expert opinion: Giuseppe Lippi

This could be an option, but I would not make it mandatory.

Editor opinion: Kaiping Zhang

No, it is not mandatory as the nature of the retrospective study. But it would undoubtedly be more evident if a mention was made.

Acknowledgments

The authors appreciate the academic support from the AME Reporting Guideline Collaborative Group. The authors thank Dr. Karel G. M. Moons (Julius Center for Health Sciences and Primary Care, UMC Utrecht, The Netherlands) for the critical comments on this study and his answers to the questions.

Funding: None.

Footnote

Data Sharing Statement: Available at <http://dx.doi.org/10.21037/atm-21-1665>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/atm-21-1665>). ZDH serves as an unpaid Executive Editor of *Annals of Translational Medicine* from Apr 2014 to Mar 2025. GL serves as an unpaid Executive Editor-in-Chief of *Annals of Translational Medicine* from Jan 2016 to Jan 2022. PMMB spearheaded the project that led to the development of the STARD reporting guidelines, published in 2003 and updated in 2015. KZ is the academic director of AME Publishing Company (the publisher of *Annals of Translational Medicine*). The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the

original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Linnet K, Bossuyt PM, Moons KG, et al. Quantifying the Accuracy of a Diagnostic Test or Marker. *Clin Chem* 2012;58:1292-301.
2. Lippi G, Plebani M. A modern and pragmatic definition of Laboratory Medicine. *Clin Chem Lab Med* 2020;58:1171.
3. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Standards for Reporting of Diagnostic Accuracy. Clin Chem* 2003;49:1-6.
4. Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem* 2003;49:7-18.
5. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016;6:e012799.
6. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. *Clin Chem* 2015;61:1446-52.
7. Pecoraro V, Banzi R, Trenti T. Quality of reporting of diagnostic test accuracy studies in medical laboratory journals. *Clin Chem Lab Med* 2016;54:e319-21.
8. Prager R, Bowdridge J, Kareemi H, et al. Adherence to the Standards for Reporting of Diagnostic Accuracy (STARD) 2015 Guidelines in Acute Point-of-Care Ultrasound Research. *JAMA Netw Open* 2020;3:e203871.
9. Hong PJ, Korevaar DA, McGrath TA, et al. Reporting of imaging diagnostic accuracy studies with focus on MRI subgroup: Adherence to STARD 2015. *J Magn Reson Imaging* 2018;47:523-44.
10. Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;3:25.
11. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529-36.
12. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol* 1991;44:763-70.
13. Korevaar DA, Gopalakrishna G, Cohen JF, et al. Targeted test evaluation: a framework for designing diagnostic accuracy studies with clear study hypotheses. *Diagn Progn Res* 2019;3:22.
14. Moons KG, Grobbee DE. When should we remain blind and when should our eyes remain open in diagnostic studies? *J Clin Epidemiol* 2002;55:633-6.
15. Peat G, Riley RD, Croft P, et al. Improving the transparency of prognosis research: the role of reporting, data sharing, registration, and protocols. *PLoS Med* 2014;11:e1001671.
16. Moons KG, Biesheuvel CJ, Grobbee DE. Test research versus diagnostic research. *Clin Chem* 2004;50:473-6.
17. van Smeden M, Naaktgeboren CA, Reitsma JB, et al. Latent class models in diagnostic studies when there is no reference standard--a systematic review. *Am J Epidemiol* 2014;179:423-31.
18. Biesheuvel CJ, Vergouwe Y, Oudega R, et al. Advantages of the nested case-control design in diagnostic research. *BMC Med Res Methodol* 2008;8:48.

Cite this article as: Zheng FF, Shen WH, Gong F, Hu ZD, Lippi G, Šimundić AM, Bossuyt PM, Plebani M, Zhang K. Adherence to the Standards for Reporting of Diagnostic Accuracy Studies (STARD): a survey of four journals in laboratory medicine. *Ann Transl Med* 2021;9(11):918. doi: 10.21037/atm-21-1665