STARD guideline in diagnostic accuracy tests: perspective from a systematic reviewer

Zhi-De Hu

Department of Laboratory Medicine, General Hospital of Ji'nan Military Region of PLA, Shandong 250031, China
Correspondence to: Zhi-De Hu. Department of Laboratory Medicine, General Hospital of Ji'nan Military Region of PLA, Shandong 250031, China.
Email: hzdlj81@163.com.

When I searched PubMed using the word "diagnostic[TI]" in December 16, 2015, I found that the number of publications has increased year by year, as shown in *Figure 1*. Notably, it increased rapidly after 2000. From 2003 [the year when the Standards for Reporting of Diagnostic Accuracy (STARD) (1) was released] to 2015, a total of 48,341 publications were retrieved. Among these publications, I screened 100 that were published in 2015 and verified 28 publications that were original articles regarding the diagnostic accuracy tests. This means, from 2003 to 2015, that there are approximately 13,500 papers regarding diagnostic accuracy tests in PubMed. Next, I used Google Scholar to search the citation times of the STARD guideline, and found that all versions of STARD guideline have been cited for 3,088 times, while in Web of Science database, all version of STARD guideline has been cited for 2,291 times.

Let's assume that authors complied with STARD recommendations will cite the STARD document when reporting their works. These results suggested that, during the past decade, the number of publications in diagnostic accuracy tests increased rapidly. However, only less than 20% of the study reports complied with STARD guideline. Indeed, some of the researchers believe that the introduction of STARD guideline did not significantly improve the quality of the reports (2), or only slightly improved it (3,4). This is an emergent condition, since nonstandard report may mislead readers to erroneously evaluate the design quality and clinical applications of the study.

Over the past few years, my colleagues and I have published some systematic reviews on diagnostic markers (5,6). When assessing the quality of available studies using Quality Assessment of Diagnostic Accuracy Studies
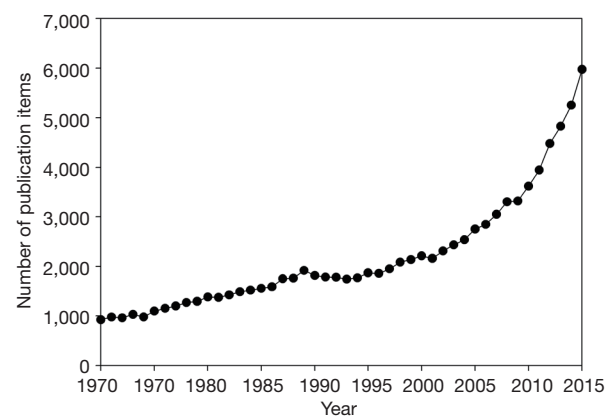

Figure 1 Number of papers retrieved from PubMed using the keyword "diagnostic[TI]".

(QUADAS) criteria (7,8), we noted that some of the authors did not report their work in accordance with the STARD guideline, and therefore, we could not estimate the risk of bias and applicability concerns for their study. For example, in studies that evaluate the diagnostic accuracy of osteopontin for ovarian cancer, some of publications did not report the exclusion and inclusion criteria for subjects enrollment, neither how the study population was enrolled (5). They only reported the disease spectrum and the sample sizes. Therefore, we did not know whether the prevalence of target disease in study cohort was consistent with the real word. The prevalence of the target disease in study cohorts is vital, because it can greatly affect estimates of test performance (9). In addition, because the inclusion and exclusion criteria were not reported, we did not know under which condition the index test should be adopted.

Recently, the STARD guideline has been updated (10).

Table 1 STARD checklist items that associated with QUADAS tools

Questions in each domain of QUADAS-2 tools	STARD items	
	2015	2003
Patients selection domain		
Was a consecutive or random sample of patients enrolled?	9	5
Was a case-control design avoided?	6, 7	4, 5
Did the study avoid inappropriate exclusions?	6, 7	4, 5
Index test domain		
Were the index test results interpreted without knowledge of the results of the reference standard?	13a	11
If a threshold was used, was it pre-specified?	12a, 17	9
Reference Standard domain		
Is the reference standard likely to correctly classify the target condition?	11, 12b	7
Were the reference standard results interpreted without knowledge of the results of the index test?	13b	11
Flow and Timing domain		
Was there an appropriate interval between index tests and reference standard?	22	17
Did all patients receive a reference standard?	19	16
Did all patients receive the same reference standard?	?	?
Were all patients included in the analysis?	19, 23	17, 19

?, Not mentioned. QUADAS, Quality Assessment of Diagnostic Accuracy Studies; STARD, Standards for Reporting of Diagnostic Accuracy.

Compared with the 2003 version, some new items were added, such as sample size estimation, registration and sources of funding. Additionally, some items have been revised or removed from the 2003 version. For example, the 10th item in the 2003 version (context bias) has been removed from the 2015 version. Generally, I believe that STARD guideline of the 2015 version will greatly improve the quality of the reports for diagnostic accuracy test and can facilitate the reader to better understand the strengths and weaknesses of a study.

It is well known that biases usually occur when the study design is defective (11). Many types of bias have been identified so far (9), such as disease progression bias, partial verification bias, differential verification bias, incorporation bias and context bias. We noticed that if an author reported the work in accordance with updated STARD guideline, most of the biases can be verified (*Table 1*). However, it seems that there is no item regarding differential verification bias listed in the updated STARD guideline. Differential verification bias is defined as part of the index test results and is verified by a different reference standard (9,11). This type of bias usually occurs when the reference standard for target disease is an invasive procedure (e.g., cancer).

I noted that an explanation and elaboration (12) for the 2003 STARD guideline has been released in association with the STARD checklist (1). However, no explanation or elaboration for the 2015 STARD checklist has been released so far. Since the items in the 2015 STARD guideline have been extensively revised from the 2003 version, a comprehensive elaboration is essential to help the authors, reviewers and journal editors to better understand the meaning, rationale and optimal use of each item on the checklist. This would be an icing on the cake, I believe.

Acknowledgements

Footnote

References

1. 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.

2. Coppus SF, van der Veen F, Bossuyt PM, et al. Quality of reporting of test accuracy studies in reproductive medicine: impact of the Standards for Reporting of Diagnostic Accuracy (STARD) initiative. Fertil Steril 2006;86:1321-9.

3. Smidt N, Rutjes AW, van der Windt DA, et al. The quality of diagnostic accuracy studies since the STARD statement: has it improved? Neurology 2006;67:792-7.

4. Korevaar DA, van Enst WA, Spijker R, et al. Reporting quality of diagnostic accuracy studies: a systematic review and meta-analysis of investigations on adherence to STARD. Evid Based Med 2014;19:47-54.

5. Hu ZD, Wei TT, Yang M, et al. Diagnostic value of osteopontin in ovarian cancer: a meta-analysis and systematic review. PLoS One 2015;10:e0126444.

6. Zhang J, Hu ZD, Song J, et al. Diagnostic value of presepsin for sepsis: a systematic review and meta-analysis. Medicine (Baltimore) 2015;94:e2158.

7. 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.

8. 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.

9. Whiting P, Rutjes AW, Reitsma JB, et al. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. Ann Intern Med 2004;140:189-202.

10. 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.

11. Schmidt RL, Factor RE. Understanding sources of bias in diagnostic accuracy studies. Arch Pathol Lab Med 2013;137:558-65.

12. 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.