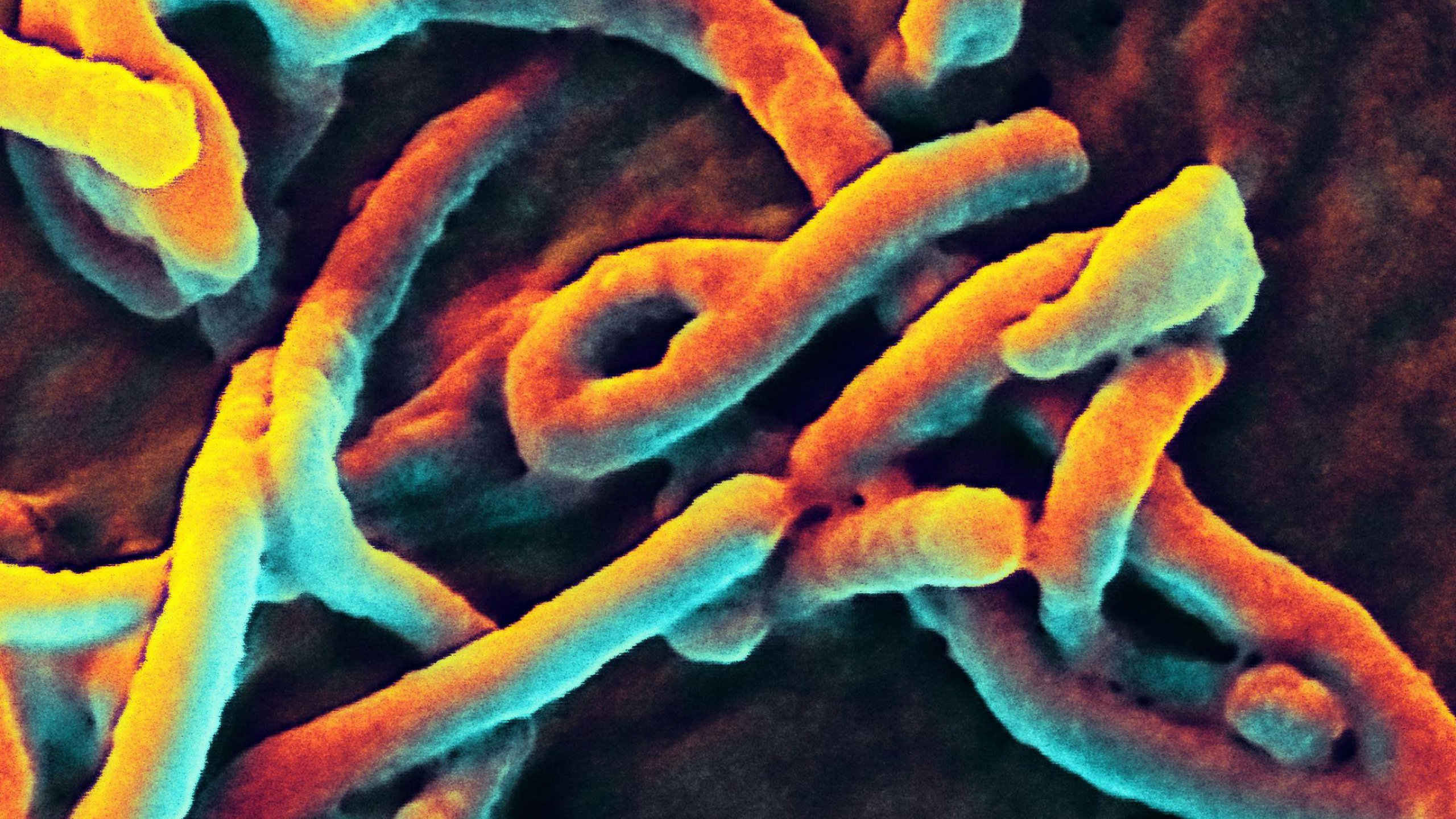
How Reporting Guidelines Can Help to Improve Practice

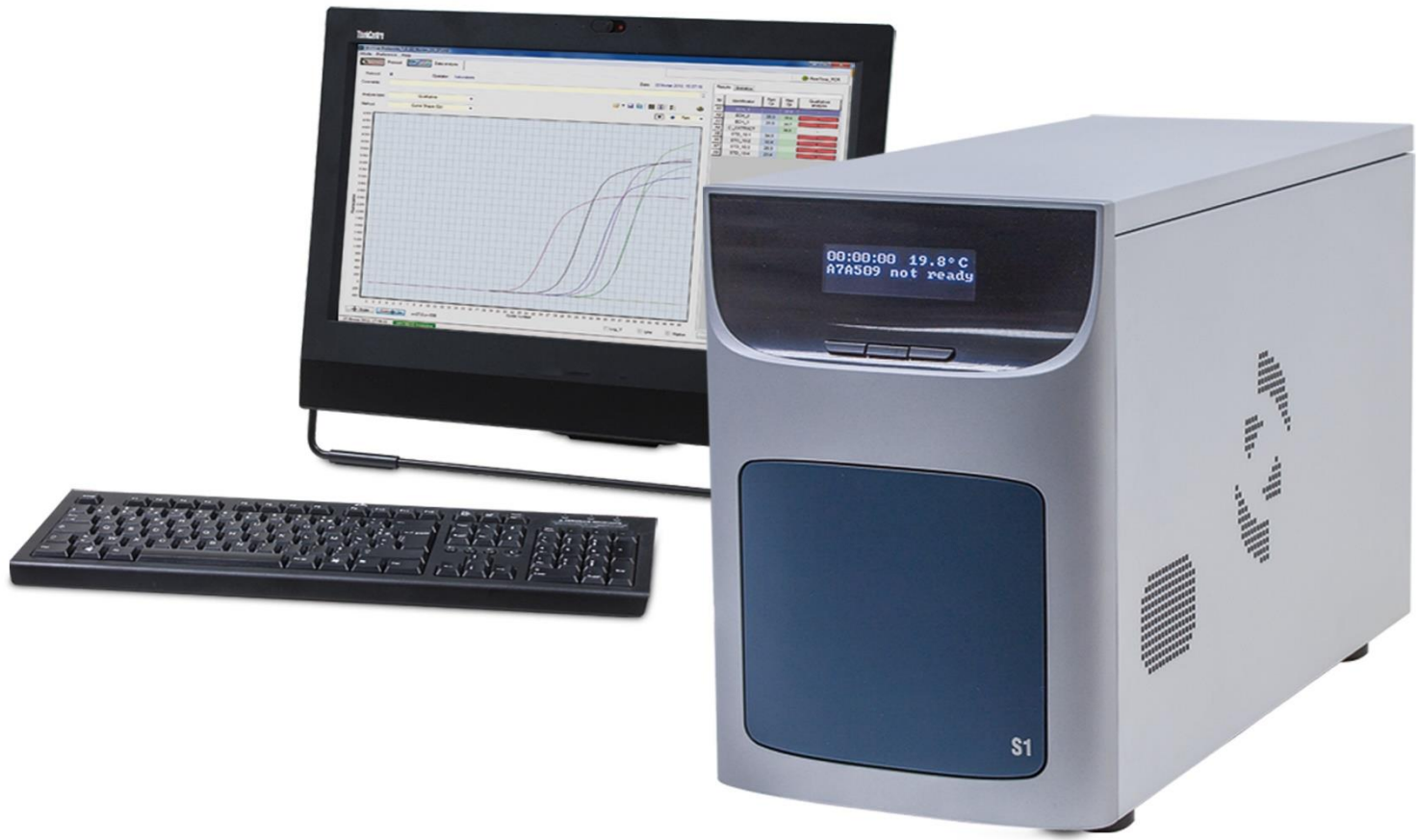
The story of STARD

Patrick M Bossuyt

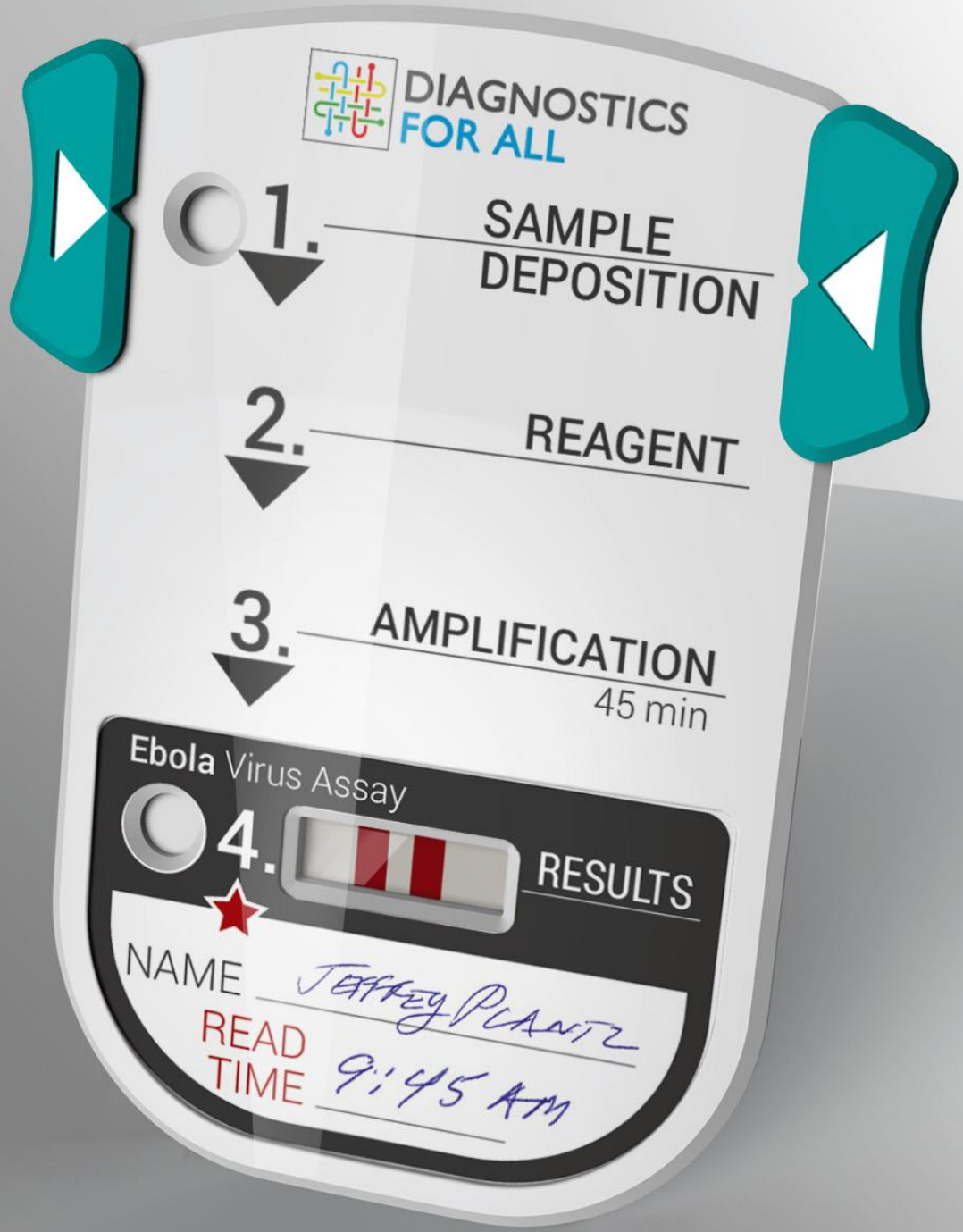
AMC – University of Amsterdam

REWARD
REduce research Waste And Reward Diligence

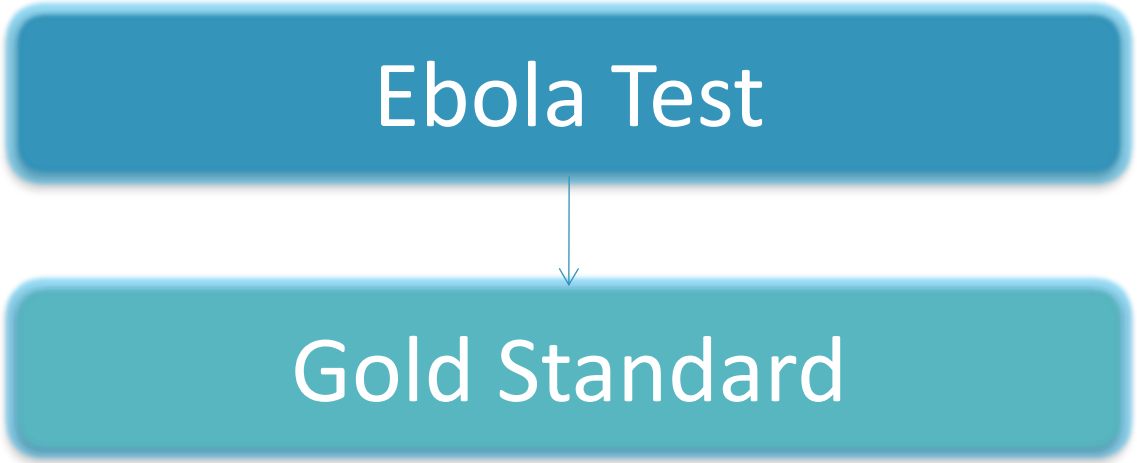



00:00:00 19.8°C
A7A509 not ready
S1

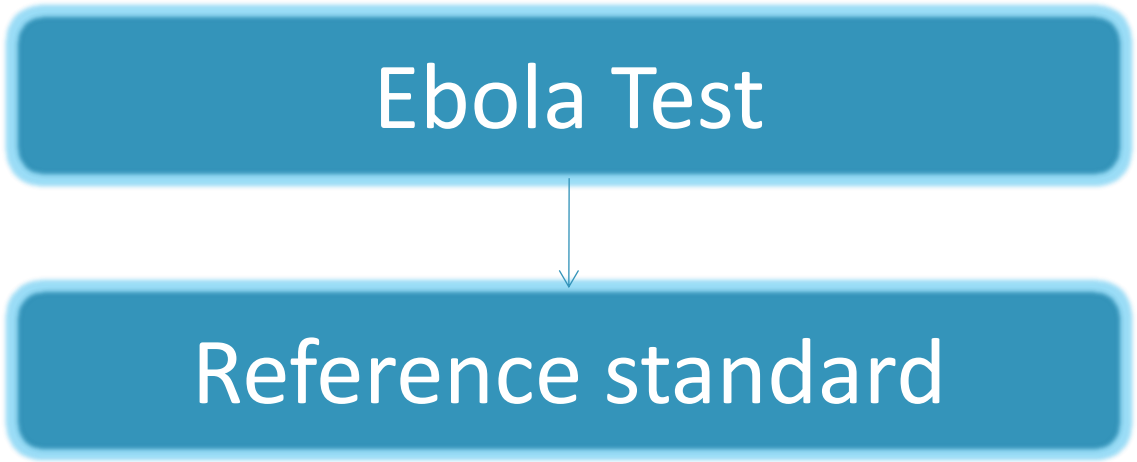

Ebola is
REAL
West Point Mon-Lib
Cell*0888


DIAGNOSTICS FOR ALL
1.
SAMPLE DEPOSITION
2.
REAGENT
3.
AMPLIFICATION
45 min
Ebola Virus Assay
4.
RESULTS
NAME
Jeffrey Plantz
READ TIME
9:45 AM

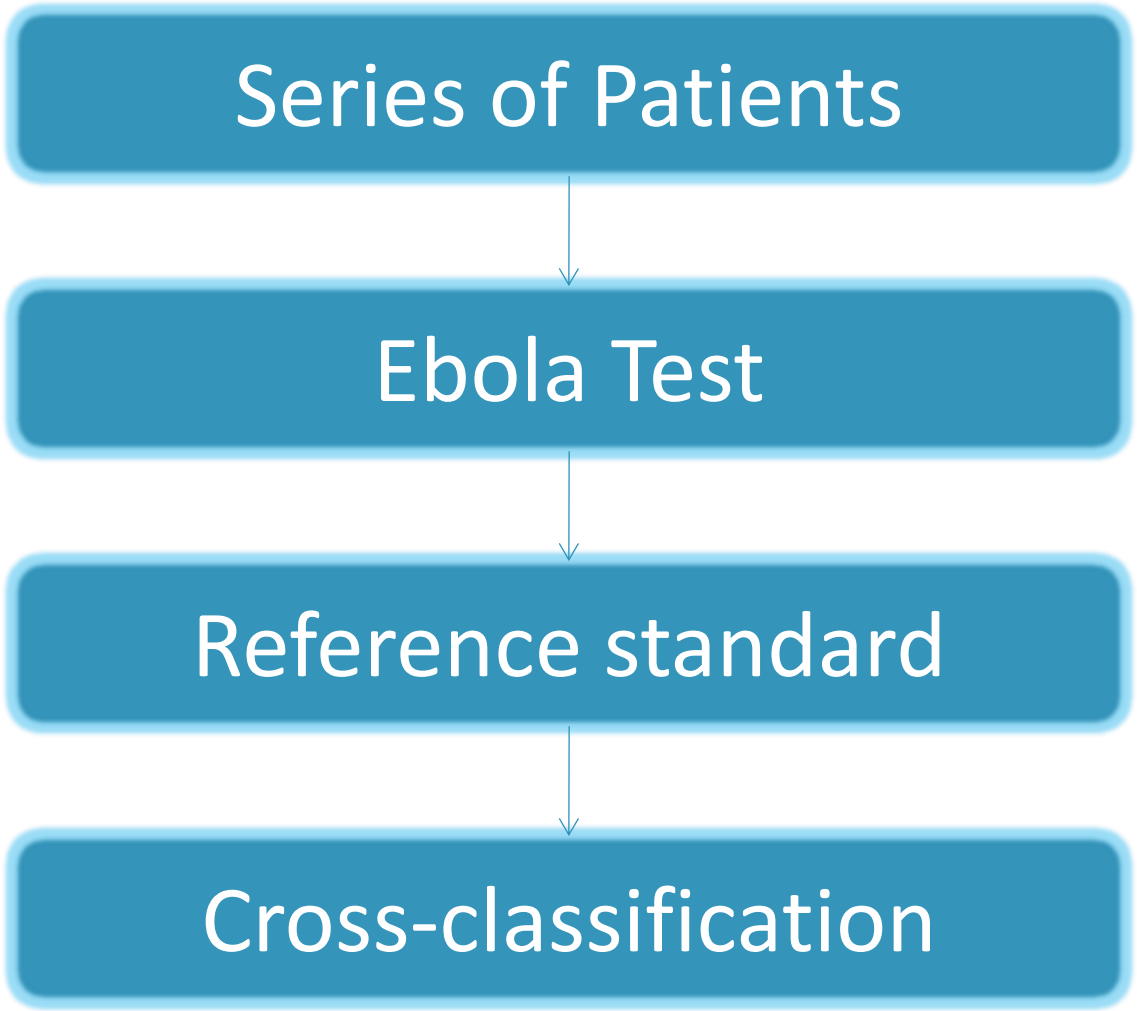
Diagnostic Accuracy



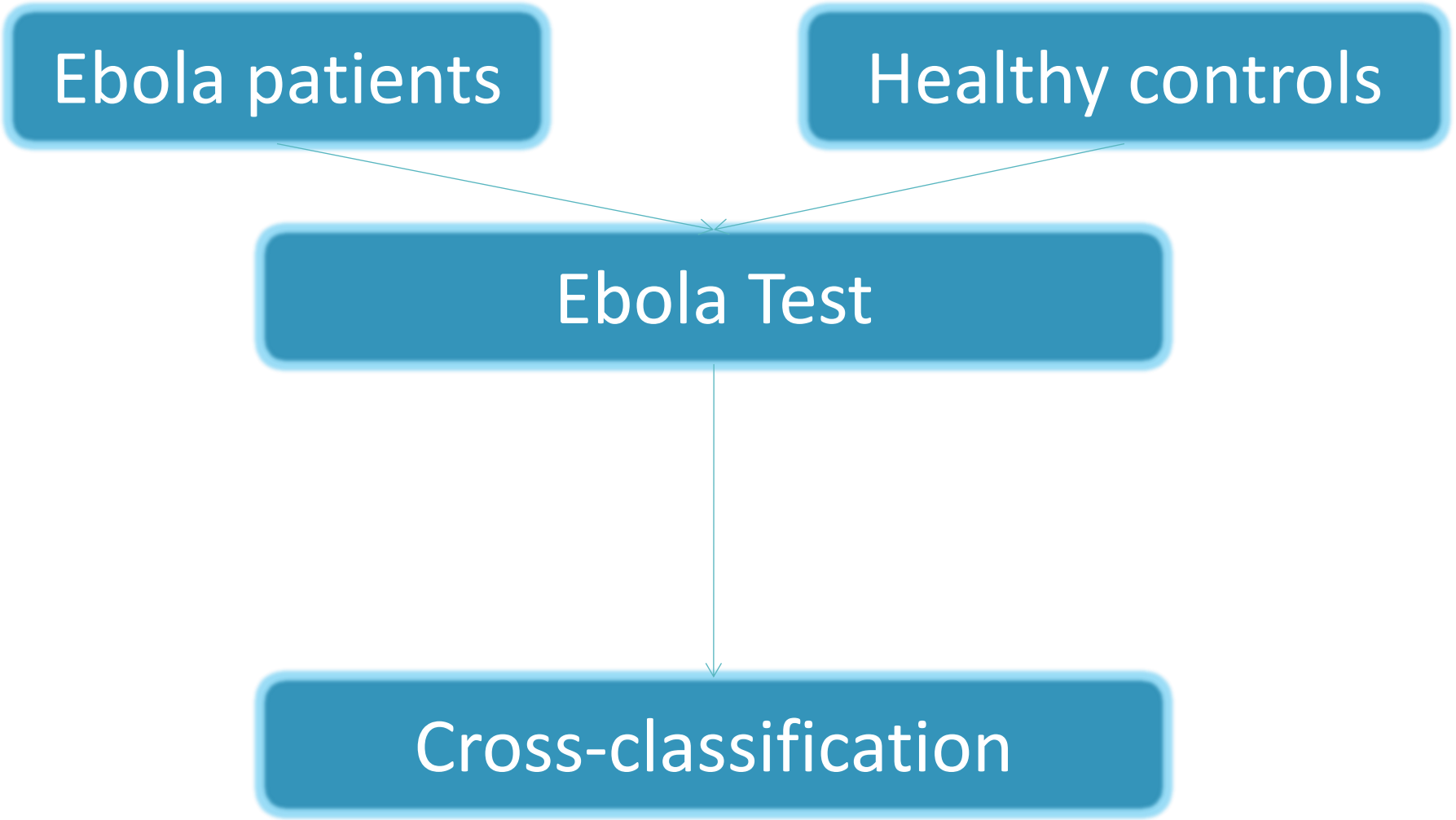
Diagnostic Accuracy



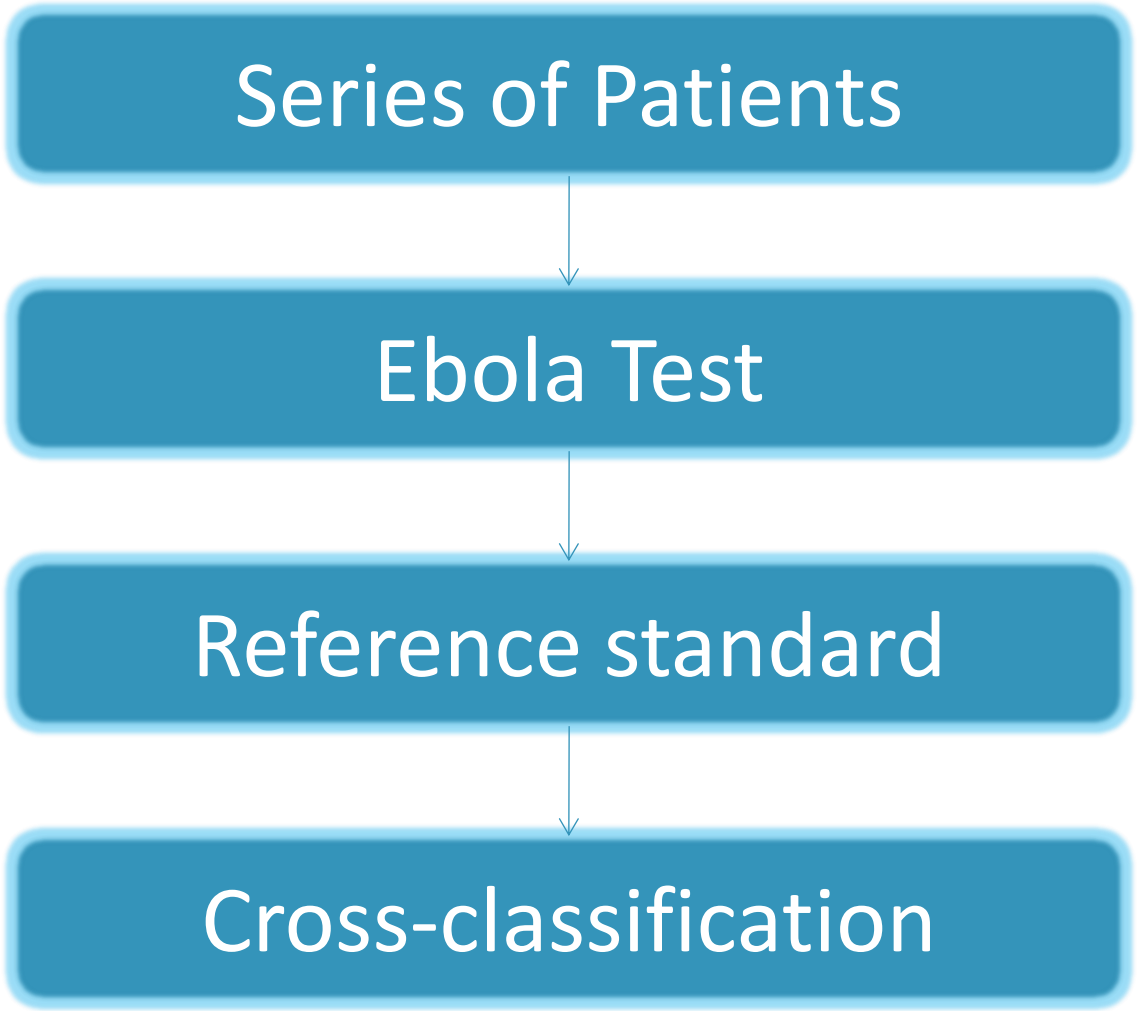
Diagnostic Accuracy Study



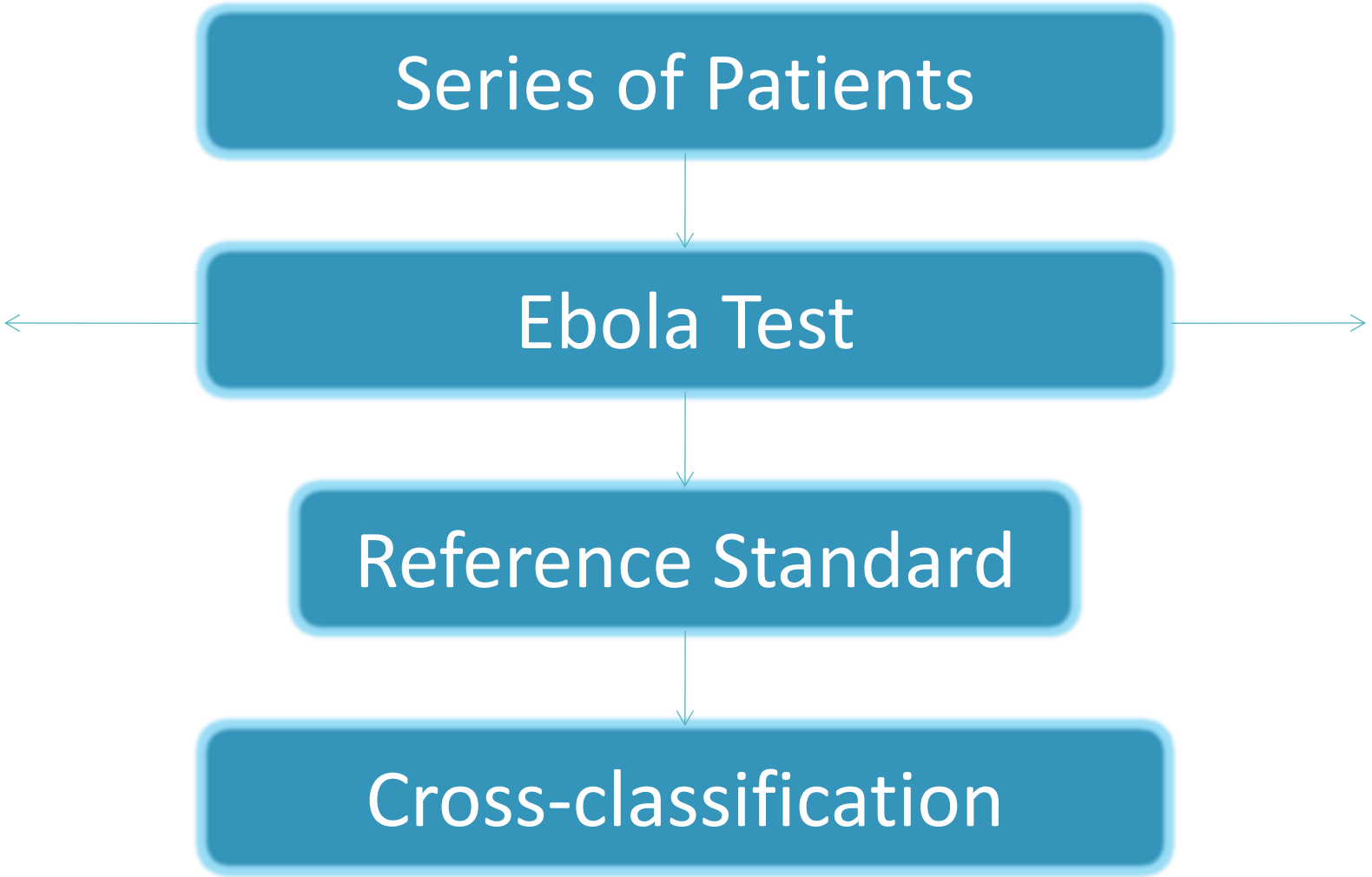
Two sets of Eligibility Criteria



Diagnostic Accuracy Study



Verification - partial



Empirical Evidence of Design-Related Bias in Studies of Diagnostic Tests

Jeroen G. Lijmer, MD
Ben Willem Mol, MD, PhD
Siem Heisterkamp, PhD
Gouke J. Bonsel, MD, PhD
Martin H. Prins, MD, PhD
Jan H. P. van der Meulen, MD, PhD
Patrick M. M. Bossuyt, PhD

Context The literature contains a large number of potential biases in the evaluation of diagnostic tests. Strict application of appropriate methodological criteria would invalidate the clinical application of most study results.

Objective To empirically determine the quantitative effect of study design shortcomings on estimates of diagnostic accuracy.

Design and Setting Observational study of the methodological features of 184 original studies evaluating 218 diagnostic tests. Meta-analyses on diagnostic tests were identified through a systematic search of the literature using MEDLINE, EMBASE, and DARE databases and the Cochrane Library (1996-1997). Associations between study characteristics and estimates of diagnostic accuracy were evaluated with a regression model.

Main Outcome Measures Relative diagnostic odds ratio (RDOR), which compared the diagnostic odds ratios of studies of a given test that lacked a particular methodological feature with those without the corresponding shortcomings in design.

Results Fifteen (6.8%) of 218 evaluations met all 8 criteria; 64 (30%) met 6 or more. Studies evaluating tests in a diseased population and a separate control group overestimated the diagnostic performance compared with studies that used a clinical population (RDOR, 3.0; 95% confidence interval [CI], 2.0-4.5). Studies in which different reference tests were used for positive and negative results of the test under study overestimated the diagnostic performance compared with studies using a single reference test for all patients (RDOR, 2.2; 95% CI, 1.5-3.3). Diagnostic performance was also overestimated when the reference test was interpreted with knowledge of the test result (RDOR, 1.3; 95% CI, 1.0-1.9), when no criteria for the test were described (RDOR, 1.7; 95% CI, 1.1-2.5), and when no description of the population under study was provided (RDOR, 1.4; 95% CI, 1.1-1.7).

Conclusion These data provide empirical evidence that diagnostic studies with methodological shortcomings may overestimate the accuracy of a diagnostic test, particularly those including nonrepresentative patients or applying different reference standards.

JAMA. 1999;282:1061-1066 www.jama.com

DURING RECENT DECADES, THE number of available diagnostic tests has been rapidly increasing. As for all new medical technologies, new diagnostic tests should be thoroughly evaluated prior to their introduction into daily practice. The number of test evaluations in the literature is increasing but the methodological quality of these studies is on average poor. A survey of the diagnostic literature (1990-1993) showed that only 18% of the studies satisfied 5 of the 7 methodological standards examined.[1] Different guidelines have been written to help physicians with the critical appraisal of the diagnostic literature consisting of lists of criteria for the assessment of study quality.[2-4] Criteria enable

JAMA 15 SEP 1999
282; 1061-1066

How often reported?

112 diagnostic accuracy studies published in 2012

Item	Reported
Inclusion and exclusion criteria	
Participant sampling: consecutive vs. random vs. convenience	
Blinding of index test readers	
Baseline characteristics (age, sex, presenting symptoms)	

Korevaar et al.
Radiology 2015
274(3):781-9

STARD: A reporting guideline

STARD (2003)

Towards Complete and Accurate Reporting of Studies of Diagnostic Accuracy: The STARD Initiative

Patrick M. Bossuyt,[1*] Johannes B. Reitsma,[1] David E. Bruns,[2,3] Constantine A. Gatsonis,[4] Paul P. Glasziou,[5] Les M. Irwig,[6] Jeroen G. Lijmer,[1] David Moher,[7] Drummond Rennie,[8,9] and Henrica C.W. de Vet,[10] for the STARD Group

Table 1. STARD checklist for the reporting of studies of diagnostic accuracy.

Section and Topic	Item #		On page #
TITLE/ABSTRACT/ KEYWORDS	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	
INTRODUCTION	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	
METHODS		Describe	
Participants	3	The study population: The inclusion and exclusion criteria, setting and locations where the data were collected.	
	4	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	
	5	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in items 3 and 4? If not, specify how participants were further selected.	
	6	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	
Test methods	7	The reference standard and its rationale.	
	8	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.	
	9	Definition of and rationale for the units, cutoffs and/or categories of the results of the index tests and the reference standard.	
	10	The number, training and expertise of the persons executing and reading the index tests and the reference standard.	
	11	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.	
Statistical methods	12	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	
	13	Methods for calculating test reproducibility, if done.	
RESULTS		Report	
Participants	14	When study was done, including beginning and ending dates of recruitment.	
	15	Clinical and demographic characteristics of the study population (e.g. age, sex, spectrum of presenting symptoms, comorbidity, current treatments, recruitment centers).	
	16	The number of participants satisfying the criteria for inclusion that did or did not undergo the index tests and/or the reference standard; describe why participants failed to receive either test (a flow diagram is strongly recommended).	
Test results	17	Time interval from the index tests to the reference standard, and any treatment administered between.	
	18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	
	19	A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	
	20	Any adverse events from performing the index tests or the reference standard.	
Estimates	21	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	
	22	How indeterminate results, missing responses and outliers of the index tests were handled.	
	23	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	
	24	Estimates of test reproducibility, if done.	
DISCUSSION	25	Discuss the clinical applicability of the study findings.	

Radiology

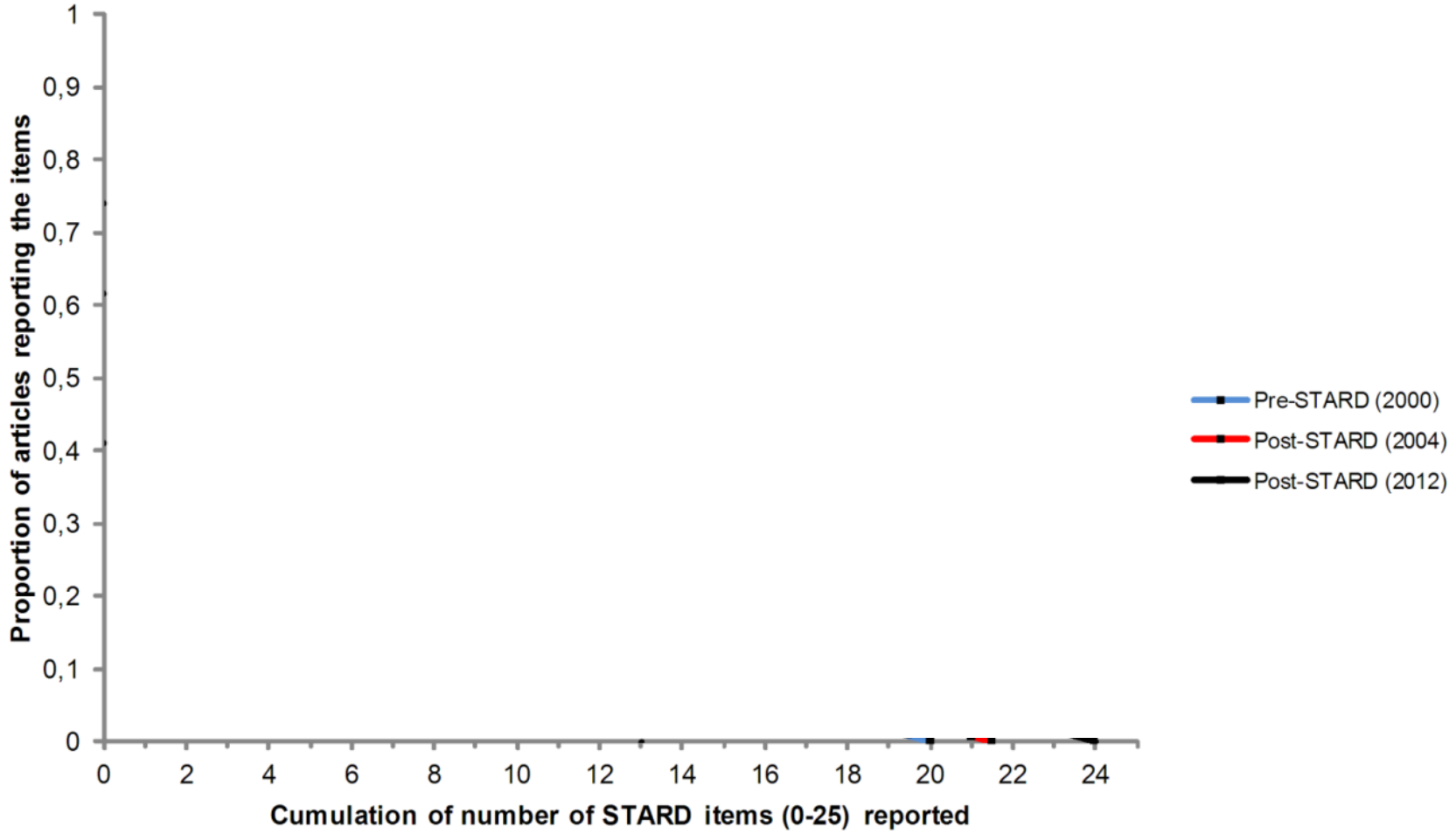
Reporting Diagnostic Accuracy Studies: Some Improvements after 10 Years of STARD[1]

Daniël A. Korevaar, MD
Junfeng Wang, MS
W. Annefloor van Enst, MS
Mariska M. Leeflang, PhD
Lotty Hooft, PhD
Nynke Smidt, PhD
Patrick M. M. Bossuyt, PhD

Purpose: To evaluate how diagnostic accuracy study reports published in 2012 adhered to the Standards for Reporting of Diagnostic Accuracy (STARD) statement and whether there were any differences in reporting compared with 2000 and 2004.

Materials and Methods: PubMed was searched for studies published in 12 high-impact-factor journals in 2012 that evaluated the accuracy of one or more diagnostic tests against a clinical reference standard. Two independent reviewers scored reporting completeness of each article with the 25-item STARD checklist. Mixed-effects modeling was used to analyze differences in reporting with previous evaluations from articles published in 2000 and 2004.

Adherence to STARD



Study or Subgroup	Post-STARD Mean	SD	Total	Pre-STARD Mean	SD	Total	Weight	Mean Difference IV, Random, 95% CI	Year
Smidt 2006	13.55	3.2	141	11.94	3.3	124	19.4%	1.61 [0.82, 2.40]	2006
Coppus 2006	12.41	3.2	27	12.08	3.3	24	10.5%	0.33 [-1.46, 2.12]	2006
Rama 2006	15.5	3.3	19	15.1	3.3	11	7.0%	0.40 [-2.05, 2.85]	2006
Wilczynski 2008*	8.43	2.22	120	7.94	2.55	120	21.1%	0.49 [-0.11, 1.09]	2008
Areia 2010	14.3	2.9	24	11.6	3.6	86	13.5%	2.70 [1.31, 4.09]	2010
Selman 2011**	12.32	2.92	17	11.05	2.57	88	12.7%	1.27 [-0.22, 2.76]	2011
Selman 2011***	15.33	3.04	35	12.75	3.54	160	15.8%	2.58 [1.43, 3.73]	2011
Total (95% CI)			383			613	100.0%	**1.41 [0.65, 2.18]**	

Heterogeneity: Tau² = 0.63; Chi² = 17.70, df = 6 (P = 0.007); I² = 66%
Test for overall effect: Z = 3.61 (P = 0.0003)

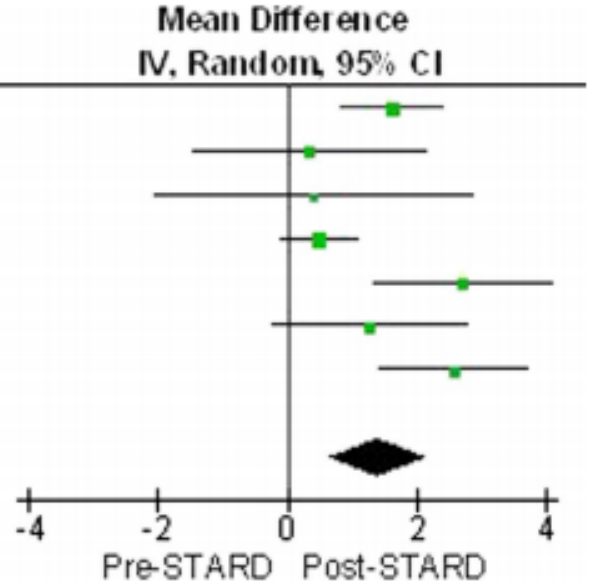


Figure 2 Forest plot for studies included in meta-analysis comparing adherence post-Standards for Reporting of Diagnostic Accuracy Studies (STARD) and pre-STARD. *Wilczynski[10] evaluated only 13 STARD items; the other studies evaluated 25 STARD items. **Results of the studies on obstetrics. ***Results of the studies on gynaecology.

Radiology

Patrick M. M. Bossuyt, PhD

STARD Statement: Still Room for Improvement in the Reporting of Diagnostic Accuracy Studies[1]

In the clinical assessment of a medical test, the evaluation of its diagnostic accuracy is an essential step. In an evaluation of diagnostic accuracy, the results of the test are compared with the results of the reference standard in the same patients. Yet, one cannot unconditionally take the results from any particular study at face value. Many authors have pointed out the multiple risks of bias in diagnostic accuracy studies (1,2). Critical appraisal of published studies is therefore essential.

Unfortunately, researchers in many published studies fail to report essential elements of study design and the checklist to evaluate the completeness of reporting and the flow diagram to simplify reporting. Editors and reviewers could use the checklist in a similar way to appraise submitted manuscripts.

After pilot testing, the STARD statement was published in January 2003 in *Radiology* and several other major scientific journals, including *BMJ*, *Clinical Chemistry*, *Lancet*, and *Annals of Internal Medicine* (6–8). A large number of journals have included the STARD statement in their instructions to authors (details on the Web site at *http://www.stard-statement.org/*).

Radiology

ORIGINAL RESEARCH ■ SPECIAL REPORT

STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies[1]

Patrick M. Bossuyt, PhD
Johannes B. Reitsma, MD, PhD
David E. Bruns, MD
Constantine A. Gatsonis, PhD
Paul P. Glasziou, MRCGP, FRACGP, PhD, MBBS
Les Irwig, MBBS, PhD
Jeroen G. Lijmer,MD, PhD
David Moher, MD, PhD
Drummond Rennie, MD, MACP, FRCP
Henrica C.W. de Vet, PhD

Incomplete reporting has been identified as a major source of avoidable waste in biomedical research. Essential information is often not provided in study reports, impeding the identification, critical appraisal, and replication of studies. To improve the quality of reporting of diagnostic accuracy studies, the Standards for Reporting of Diagnostic Accuracy Studies (STARD) statement was developed. Here we present STARD 2015, an updated list of 30 essential items that should be included in every report of a diagnostic accuracy study. This update incorpo-

Table 1

The STARD 2015 List

Section and Topic	No.	Item
TITLE OR ABSTRACT		
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)
ABSTRACT		
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)
INTRODUCTION		
	3	Scientific and clinical background, including the intended use and clinical role of the index test
	4	Study objectives and hypotheses
METHODS		
Study design	5	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)
Participants	6	Eligibility criteria
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)
	8	Where and when potentially eligible participants were identified (setting, location and dates)
	9	Whether participants formed a consecutive, random or convenience series
Test methods	10a	Index test, in sufficient detail to allow replication
	10b	Reference standard, in sufficient detail to allow replication
	11	Rationale for choosing the reference standard (if alternatives exist)
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test
	13b	Whether clinical information and index test results were available to the assessors of the reference standard
Analysis	14	Methods for estimating or comparing measures of diagnostic accuracy
	15	How indeterminate index test or reference standard results were handled
	16	How missing data on the index test and reference standard were handled
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory
	18	Intended sample size and how it was determined
RESULTS		
Participants	19	Flow of participants, using a diagram
	20	Baseline demographic and clinical characteristics of participants
	21a	Distribution of severity of disease in those with the target condition
	21b	Distribution of alternative diagnoses in those without the target condition
	22	Time interval and any clinical interventions between index test and reference standard
Test results	23	Cross tabulation of the index test results (or their distribution) by the results of the reference standard
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)
	25	Any adverse events from performing the index test or the reference standard
DISCUSSION		
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability
	27	Implications for practice, including the intended use and clinical role of the index test
OTHER INFORMATION		
	28	Registration number and name of registry
	29	Where the full study protocol can be accessed
	30	Sources of funding and other support; role of funders

STARD: Who benefits

- Authors
- Reviewers
- Editors

- Students
- Funders

- Patients

STARD: Improve terminology

Evaluation of a point-of-care blood test for identification of Ebola virus disease at Ebola holding units, Western Area, Sierra Leone, January to February 2015

N F Walker[1,2,3], C S Brown[1,4], D Youkee[1], P Baker[1], N Williams[5], A Kalawa[5], K Russell[6], A F Samba[7], N Bentley[6], F Koroma[7], M B King[7], B E Parker[7], M Thompson[7], T Boyles[1,8], B Healey[1], B Kargbo[7], D Bash-Taqi[7], A J Simpson[6], A Kamara[7], T B Kamara[7], M Lado[1], O Johnson[1], T Brooks (Tim.Brooks@phe.gov.uk)[6]

1. King's Sierra Leone Partnership, King's Centre for Global Health, King's College London, and King's Health Partners, London, United Kingdom
2. Department of Infectious Diseases and Immunity, Imperial College London, London, United Kingdom
3. Clinical Infectious Diseases Research Initiative, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa
4. Hospital for Tropical Diseases, University College London Hospital, London, United Kingdom
5. College of Medicine and Allied Health Sciences, Freetown, Sierra Leone
6. Rare and Imported Pathogens Laboratory, Public Health England, Salisbury, United Kingdom
7. Ministry of Health and Sanitation, Sierra Leone
8. Division of Infectious Diseases and HIV Medicine, Department of Medicine, University of Cape Town, Cape Town, South Africa

Citation style for this article:
Walker NF, Brown CS, Youkee D, Baker P, Williams N, Kalawa A, Russell K, Samba AF, Bentley N, Koroma F, King MB, Parker BE, Thompson M, Boyles T, Healey B, Kargbo B, Bash-Taqi D, Simpson AJ, Kamara A, Kamara TB, Lado M, Johnson O, Brooks T. Evaluation of a point-of-care blood test for identification of Ebola virus disease at Ebola holding units, Western Area, Sierra Leone, January to February 2015. Euro Surveill. 2015;20(12):pii=21073. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=21073

Article submitted on 16 March 2015 / published on 26 March 2015

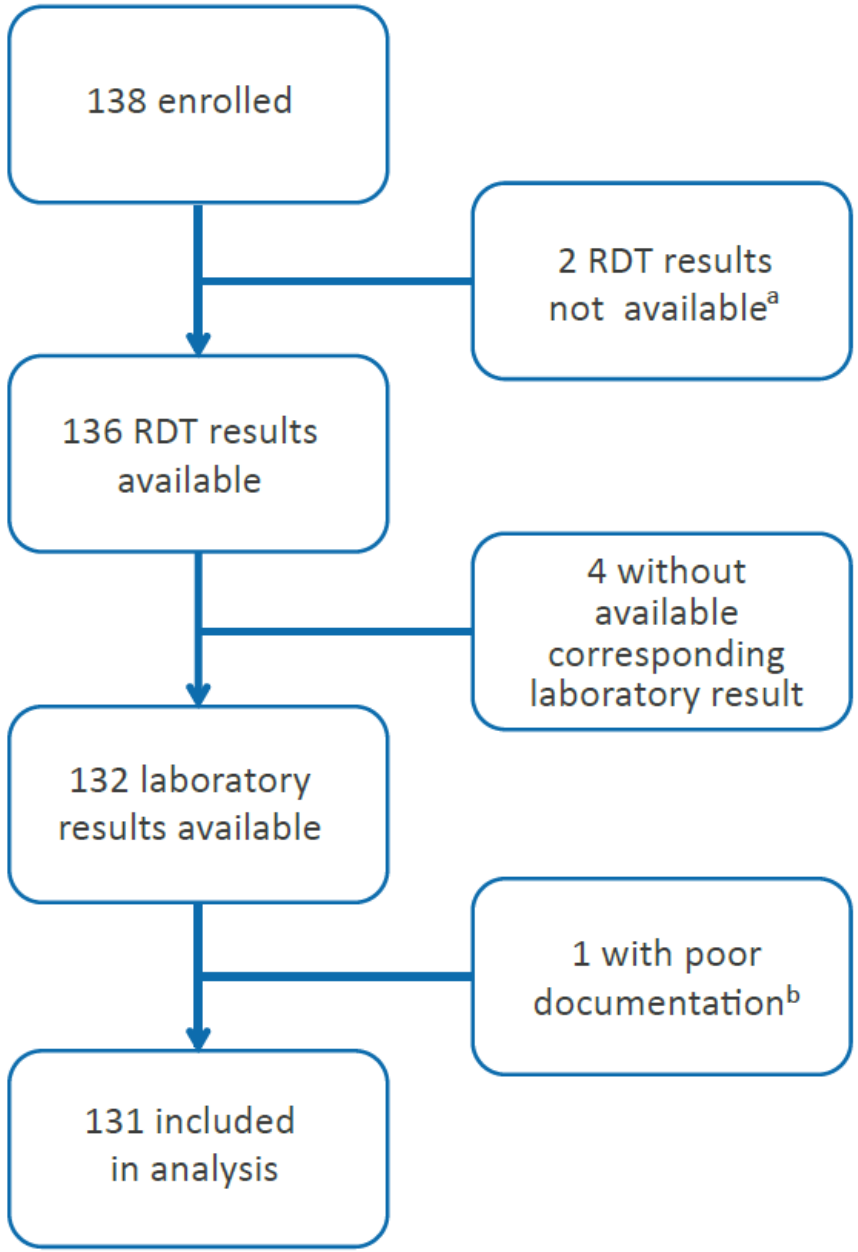


FIGURE 1

DSTL rapid diagnostic antigen test for Ebola virus disease, study enrolment and inclusion, Sierra Leone, January–February 2015 (n = 138)

DSTL: Defence Science and Technology Laboratory; RDT: rapid diagnostic antigen test.

[a] In one case the RDT attempt failed as an extremely small volume of blood was collected after the pinprick, in a second case no RDT result was documented.

[b] Possible double entry of a patient with discordant RDT results

Current Ebola virus disease (EVD) diagnosis relies on reverse transcription-PCR (RT-PCR) technology, requiring skilled laboratory personnel and technical infrastructure. Lack of laboratory diagnostic capacity has led to diagnostic delays in the current West African EVD outbreak of 2014 and 2015, compromising outbreak control. We evaluated the diagnostic accuracy of the EVD bedside rapid diagnostic antigen test (RDT) developed by the United Kingdom's Defence Science and Technology Laboratory, compared with Ebola virus RT-PCR, in an operational setting for EVD diagnosis of suspected cases admitted to Ebola holding units in the Western Area of Sierra Leone. From 22 January to 16 February 2015, 138 participants were enrolled. EVD prevalence was 11.5%. All EVD cases were identified by a positive RDT with a test line score of 6 or more, giving a sensitivity of 100% (95% confidence interval (CI): 78.2–100). The corresponding specificity was high (96.6%, 95% CI: 91.3–99.1). The positive and

The Altona assay has been selected by the World Health Organization (WHO) as the reference standard for this outbreak.

STARD: Improve standards

Annals of Internal Medicine | Research and Reporting Methods

QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies

Penny F. Whiting, PhD; Anne W.S. Rutjes, PhD; Marie E. Westwood, PhD; Susan Mallett, PhD; Jonathan J. Deeks, PhD; Johannes B. Reitsma, MD, PhD; Mariska M.G. Leeflang, PhD; Jonathan A.C. Sterne, PhD; Patrick M.M. Bossuyt, PhD; and the QUADAS-2 Group*

In 2003, the QUADAS tool for systematic reviews of diagnostic accuracy studies was developed. Experience, anecdotal reports, and feedback suggested areas for improvement; therefore, QUADAS-2 was developed. This tool comprises 4 domains: patient selection, index test, reference standard, and flow and timing. Each domain is assessed in terms of risk of bias, and the first 3 domains are also assessed in terms of concerns regarding applicability. Signalling questions are included to help judge risk of bias.

The QUADAS-2 tool is applied in 4 phases: summarize the review question, tailor the tool and produce review-specific guidance, construct a flow diagram for the primary study, and judge bias and applicability. This tool will allow for more transparent rating of bias and applicability of primary diagnostic accuracy studies.

Ann Intern Med. 2011;155:529-536. www.annals.org

For author affiliations, see end of text.

* For members of the QUADAS-2 Group, see the **Appendix** (available at www.annals.org).

STARD: Reducing waste

STARD and Research Ethics

“This study clearly represents a considerable investment
on the parts of not only the investigators but also the participants.

Ethical considerations oblige all of us
to ensure accurate reporting of such a study.

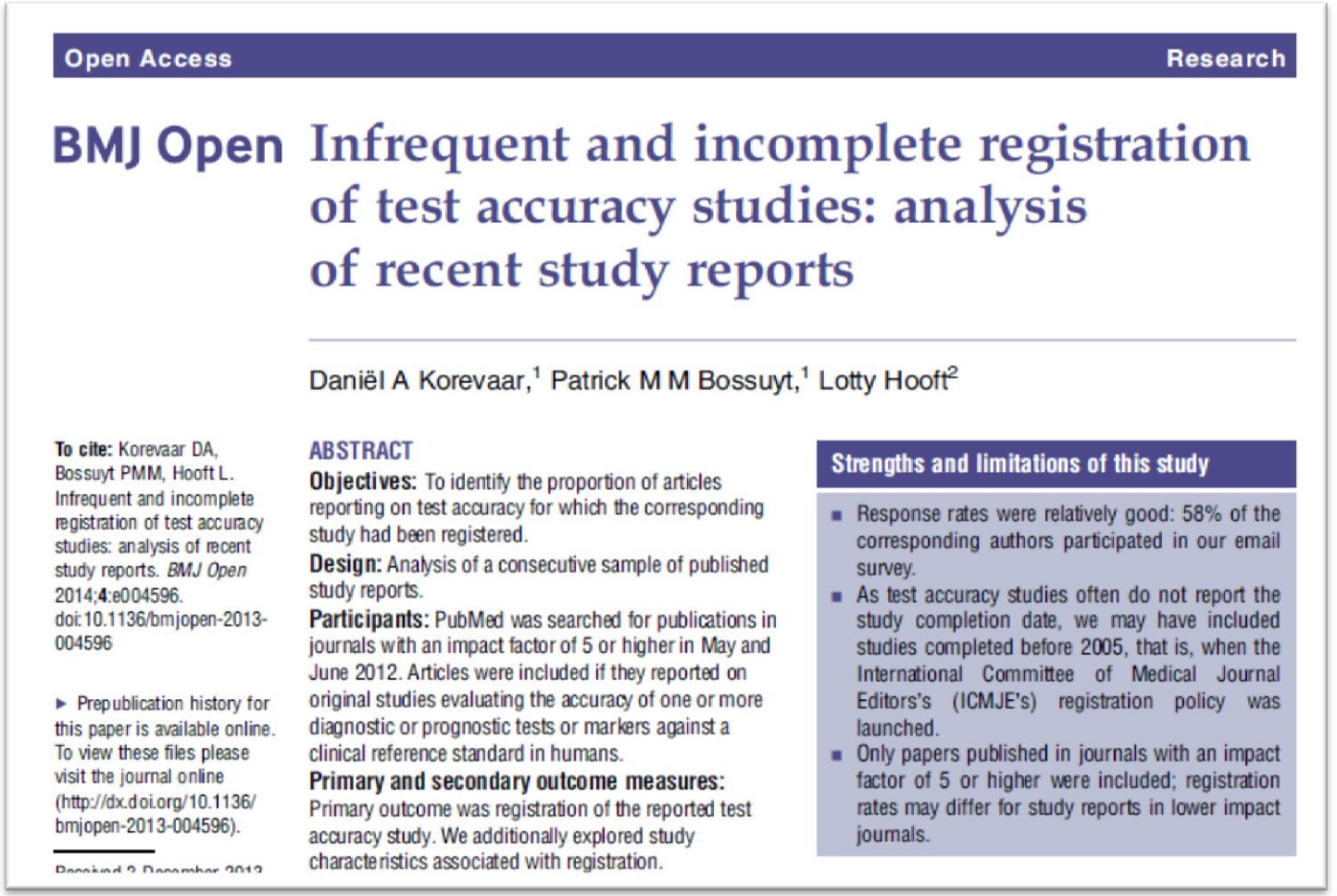
As an aid to full and clear description of the study and its results,
please consult the STARD document while revising the manuscript,
and return the completed STARD checklist with the revised
manuscript (as requested in our Information for Authors). “

STARD: A consortium to improve practice

STARD for Abstracts

Section	Item
	Identify abstract as a report of a diagnostic accuracy study (using at least one measure of accuracy, such as sensitivity, specificity, predictive values, or area under the ROC curve)
BACKGROUND	Describe: Study objectives
METHODS	Data collection: whether this is a prospective or retrospective study Eligibility criteria for participants and the settings where the data were collected Whether participants formed a consecutive, random or convenience series Description of the index test and reference standard
RESULTS	Number of participants with and without the target condition included in the analysis Estimates of accuracy with measures of statistical uncertainty
DISCUSSION	General interpretation of the results Implications for practice, including the intended use of the index test

STARD for Trial Registration



Open Access Research

BMJ Open

Infrequent and incomplete registration of test accuracy studies: analysis of recent study reports

Daniël A Korevaar,[1] Patrick M M Bossuyt,[1] Lotty Hooft[2]

To cite: Korevaar DA, Bossuyt PMM, Hooft L. Infrequent and incomplete registration of test accuracy studies: analysis of recent study reports. *BMJ Open* 2014;**4**:e004596. doi:10.1136/bmjopen-2013-004596

► Prepublication history for this paper is available online. To view these files please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2013-004596).

Received 2 December 2013

ABSTRACT

Objectives: To identify the proportion of articles reporting on test accuracy for which the corresponding study had been registered.

Design: Analysis of a consecutive sample of published study reports.

Participants: PubMed was searched for publications in journals with an impact factor of 5 or higher in May and June 2012. Articles were included if they reported on original studies evaluating the accuracy of one or more diagnostic or prognostic tests or markers against a clinical reference standard in humans.

Primary and secondary outcome measures: Primary outcome was registration of the reported test accuracy study. We additionally explored study characteristics associated with registration.

Strengths and limitations of this study

- Response rates were relatively good: 58% of the corresponding authors participated in our email survey.
- As test accuracy studies often do not report the study completion date, we may have included studies completed before 2005, that is, when the International Committee of Medical Journal Editors's (ICMJE's) registration policy was launched.
- Only papers published in journals with an impact factor of 5 or higher were included; registration rates may differ for study reports in lower impact journals.

BMJ Open

STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration

Jérémie F Cohen,[1,2] Daniël A Korevaar,[1] Douglas G Altman,[3] David E Bruns,[4] Constantine A Gatsonis,[5] Lotty Hooft,[6] Les Irwig,[7] Deborah Levine,[8,9] Johannes B Reitsma,[10] Henrica C W de Vet,[11] Patrick M M Bossuyt[1]

To cite: Cohen JF, Korevaar DA, Altman DG, *et al*. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016;**6**:e012799. doi:10.1136/bmjopen-2016-012799

► Prepublication history for this paper is available online. To view these files please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2016-012799).

JFC and DAK contributed equally to this manuscript and share first authorship.

Received 26 May 2016
Revised 3 August 2016
Accepted 25 August 2016

For numbered affiliations see end of article.

Correspondence to
Professor Patrick M M Bossuyt; p.m.bossuyt@amc.uva.nl

ABSTRACT

Diagnostic accuracy studies are, like other clinical studies, at risk of bias due to shortcomings in design and conduct, and the results of a diagnostic accuracy study may not apply to other patient groups and settings. Readers of study reports need to be informed about study design and conduct, in sufficient detail to judge the trustworthiness and applicability of the study findings. The STARD statement (Standards for Reporting of Diagnostic Accuracy Studies) was developed to improve the completeness and transparency of reports of diagnostic accuracy studies. STARD contains a list of essential items that can be used as a checklist, by authors, reviewers and other readers, to ensure that a report of a diagnostic accuracy study contains the necessary information. STARD was recently updated. All updated STARD materials, including the checklist, are available at http://www.equator-network.org/reporting-guidelines/stard. Here, we present the STARD 2015 explanation and elaboration document. Through commented examples of appropriate reporting, we clarify the rationale for each of the 30 items on the STARD 2015 checklist, and describe what is expected from authors in developing sufficiently informative study reports.

INTRODUCTION

Diagnostic accuracy studies are at risk of bias, not unlike other clinical studies. Major sources of bias originate in methodological deficiencies, in participant recruitment, data collection, executing or interpreting the test or in data analysis.[1 2] As a result, the estimates of sensitivity and specificity of the test that is compared against the reference standard can be flawed, deviating systematically from what would be obtained in ideal circumstances (see key terminology in table 1). Biased results can lead to improper recommendations about testing, negatively affecting patient outcomes or healthcare policy.

Diagnostic accuracy is not a fixed property of a test. A test's accuracy in identifying patients with the target condition typically varies between settings, patient groups and depending on prior testing.[2] These sources of variation in diagnostic accuracy are relevant for those who want to apply the findings of a diagnostic accuracy study to answer a specific question about adopting the test in his or her environment. Risk of bias and concerns about the applicability are the two key components of QUADAS-2, a quality assessment tool for diagnostic accuracy studies.[3]

Readers can only judge the risk of bias and applicability of a diagnostic accuracy study if they find the necessary information to do so in the study report. The published study report has to contain all the essential information to judge the trustworthiness and relevance of the study findings, in addition to a complete and informative disclose about the study results.

Unfortunately, several surveys have shown that diagnostic accuracy study reports often fail to transparently describe core elements.[4–6] Essential information about included patients, study design and the actual results is frequently missing, and recommendations about the test under evaluation are often generous and too optimistic.

To facilitate more complete and transparent reporting of diagnostic accuracy studies, the STARD statement was developed: Standards for Reporting of Diagnostic Accuracy Studies.[7] Inspired by the Consolidated Standards for the Reporting of Trials or CONSORT statement for reporting randomised controlled trials,[8 9] STARD contains a checklist of items that should be reported in any diagnostic accuracy study.

The STARD statement was initially released in 2003 and updated in 2015.[10] The objectives of this update were to include recent evidence about sources of bias and variability and other issues in complete reporting, and

Downloaded from http://bmjopen.bmj.com/ on November 27, 2016 - Published by group.bmj.com

Table 1 Key STARD terminology

Term	Explanation
Medical test	Any method for collecting additional information about the current or future health status of a patient
Index test	The test under evaluation
Target condition	The disease or condition that the index test is expected to detect
Clinical reference standard	The best available method for establishing the presence or absence of the target condition. A gold standard would be an error-free reference standard
Sensitivity	Proportion of those with the target condition who test positive with the index test
Specificity	Proportion of those without the target condition who test negative with the index test
Intended use of the test	Whether the index test is used for diagnosis, screening, staging, monitoring, surveillance, prediction, prognosis or other reasons
Role of the test	The position of the index test relative to other tests for the same condition (eg, triage, replacement, add-on, new test)
Indeterminate results	Results that are neither positive or negative

make the STARD list easier to use. The updated STARD 2015 list now has 30 essential items (table 2).

Below, we present an explanation and elaboration of STARD 2015. This is an extensive revision and update of a similar document that was prepared for the STARD 2003 version.[11] Through commented examples of appropriate reporting, we clarify the rationale for each item and describe what is expected from authors.

We are confident that these descriptions can further assist scientists in writing fully informative study reports, and help peer reviewers, editors and other readers in verifying that submitted and published manuscripts of diagnostic accuracy studies are sufficiently detailed.

STARD 2015 ITEMS: EXPLANATION AND ELABORATION

Title or abstract

Item 1. Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values or AUC)

Example. 'Main outcome measures: Sensitivity and specificity of CT colonography in detecting individuals with advanced neoplasia (i.e., advanced adenoma or colorectal cancer) 6 mm or larger'.[12]

Explanation. When searching for relevant biomedical studies on a certain topic, electronic databases such as MEDLINE and Embase are indispensable. To facilitate retrieval of their article, authors can explicitly identify it as a report of a diagnostic accuracy study. This can be performed by using terms in the title and/or abstract that refer to measures of diagnostic accuracy, such as 'sensitivity', 'specificity', 'positive predictive value', 'negative predictive value', 'area under the ROC curve (AUC)' or 'likelihood ratio'.

In 1991, MEDLINE introduced a specific keyword (MeSH heading) for indexing diagnostic studies: 'Sensitivity and Specificity.' Unfortunately, the sensitivity of using this particular MeSH heading to identify diagnostic accuracy studies can be as low as 51%.[13] As of May 2015, Embase's thesaurus (Emtree) has 38 check tags for study types; 'diagnostic test accuracy study' is one of them, but was only introduced in 2011.

In the example, the authors mentioned the terms 'sensitivity' and 'specificity' in the abstract. The article will now be retrieved when using one of these terms in a search strategy, and will be easily identifiable as one describing a diagnostic accuracy study.

Abstract

Item 2. Structured summary of study design, methods, results and conclusions (for specific guidance, see STARD for Abstracts)

Example. See STARD for Abstracts (*manuscript in preparation; checklist will be available at* http://www.equator-network.org/reporting-guidelines/stard/).

Explanation. Readers use abstracts to decide whether they should retrieve the full study report and invest time in reading it. In cases where access to the full study report cannot be obtained or where time is limited, it is conceivable that clinical decisions are based on the information provided in abstracts only.

In two recent literature surveys, abstracts of diagnostic accuracy studies published in high-impact journals or presented at an international scientific conference were found insufficiently informative, because key information about the research question, study methods, study results and the implications of findings were frequently missing.[14 15]

Informative abstracts help readers to quickly appraise critical elements of study validity (risk of bias) and applicability of study findings to their clinical setting (generalisability). Structured abstracts, with separate headings for objectives, methods, results and interpretation, allow readers to find essential information more easily.[16]

Building on STARD 2015, the newly developed STARD for Abstracts provides a list of essential items that should be included in journal and conference abstracts of diagnostic accuracy studies *(list finalised; manuscript under development)*.

Introduction

Item 3. Scientific and clinical background, including the intended use and clinical role of the index test

A checklist is not the end product

- A list of reporting items is only the beginning

- We have to develop real tools
  - Teaching material
  - Writing aids
  - Reviewing tools
  - …

STARD

Complete And Transparent Reporting
Of Diagnostic Accuracy Studies